



PHD

Nonlinear Time Series Analysis in Financial Applications

Miao, Robin

Award date:
2012

Awarding institution:
University of Bath

[Link to publication](#)

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

Take down policy

If you consider content within Bath's Research Portal to be in breach of UK law, please contact: openaccess@bath.ac.uk with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

Nonlinear Time Series Analysis in Financial Applications

RongHui Miao

A thesis submitted for the degree of Doctor of Philosophy

Department of Economics

University of Bath

March 2012

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with its author. A copy of this thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and they must not copy it or use material from it except as permitted by law or with the consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

Acknowledgements

I would like to take this opportunity to express my gratitude to the people who had been helpful and supportive during my study.

First and foremost, I am enormously grateful to my supervisor, Professor Christos Ioannidis, for his patient guidance, invaluable suggestions, and continuous encouragement. Without him, this thesis would not have been possible, and his feedback contributed greatly to this thesis. I thank him also for his faith in my research ability during these years. I remember our many stimulating and constructive discussions about interesting research topics in finance and other subjects. I have also learnt from him his rigorous approach to academic research.

I am also grateful to Dr Julian Williams, who is now a Lecturer in Finance at the University of Aberdeen Business School. I have learnt from Julian that innovative ideas and critical thinking are essential for becoming a good researcher. I enjoyed very much when I was working with Julian and Christos on a chapter about interest rate models that has been published in *"Optimization and Its Applications, in Handbook of Financial Engineering, Series: Springer Optimization and Its Applications Vol. 18, 2008"*.

I would also like to thank Dr Giovanni Calice, who is now a Lecturer in Finance at the University of Southampton Management School, for providing the data that are used in the last chapter of this thesis. I enjoyed working with Giovanni and Christos on our paper based on this chapter. Furthermore, Giovanni was always available for discussing interesting research topics. His research enthusiasm about the effect of credit risks on financial market's stability has also inspired me a lot in this research area.

I would like to thank my fellow doctoral students for their constructive discussions on various issues in research. Special thanks go to Xinyang Li, my former fellow at the School of Management, for her helpful advice, both academically and personally.

Last, but never least, I thank my parents for patience and precious support during my study.

Contribution of Authors

The last chapter of this thesis is a collaboration between Rong Hui Miao, Christos Ioannidis and Giovanni Calice. Christos Ioannidis is a Professor of Finance and Economics at the Department of Economics, University of Bath. Giovanni Calice is a PhD student at the Department of Economics, University of Bath and also a Lecturer in Finance at the School of Management, University of Southampton. Rong Hui Miao performed the data analysis, econometric modelling and empirical results that make up the chapter. Rong Hui Miao, Professor Christos Ioannidis and Dr. Giovanni Calice jointly developed the main research ideas. Rong Hui Miao and Professor Christos Ioannidis discussed revisions of the chapter. Professor Christos Ioannidis has contributed to the empirical results and conclusion sections of this chapter (about 20% in writing). Dr. Giovanni Calice supplied the data and contributed to the literature review of this chapter (about 10% in writing).

Abstract

The purpose of this thesis is to examine the nonlinear relationships between financial (and economic) variables within the field of financial econometrics. The thesis comprises two reviews of literatures, one on nonlinear time series models and the other one on term structure of interest rates, and four empirical essays on financial applications using nonlinear modelling techniques.

The first empirical essay compares different model specifications of a Markov switching CIR model on the term structure of UK interest rates. We find the least restricted model provides the best in-sample estimation results. Although models with restrictive specifications may provide slightly better out-of-sample forecasts in directional movements of the yields, the economic gains seem to be small.

In the second essay, we jointly model the nominal and real term structure of the UK interest rates using a three-factor essentially affine no-arbitrage term structure model. The model-implied expected inflation rates are then used in the subsequent analysis on its nonlinear relationship with the FTSE 100 index return premiums, utilizing a smooth transition vector autoregressive model. We find the model implied expected inflation rates remain below the actual inflation rates after the independence of the Bank of England in 1997, and the recent sharp decline of the expected inflation rates may lend support to the standing ground of the central bank for keeping interest rates low. The nonlinearity test on the relationship between the FTSE 100 index return premiums and the expected inflation rates shows that there exists a nonlinear adjustment on the impact from lagged expected inflation rates to current return premiums.

The third essay provides us additional insight into the nature of the aggregate stock market volatilities and its relationship to the expected returns, in a Markov switching model framework, using centuries-long aggregate stock market data from six countries (Australia, Canada, Sweden, Switzerland, UK and US). We find that the Markov switching model assuming both regime dependent mean and volatility with a 3-regime specification is capable to captures the extreme movements of the stock market which are short-lived. The volatility feedback effect that we studied on each of these six countries shows a positive sign on anticipating a high volatility regime of the

current trading month. The investigation on the coherence in regimes over time for the six countries shows different results for different pairs of countries.

In the last essay, we decompose the term premium of the North American CDX investment grade index into a permanent and a stationary component using a Markov switching unobserved component model. We explain the evolution of the two components in relating them to monetary policy and stock market variables. We establish that the inversion of the CDX index term premium is induced by sudden changes in the unobserved stationary component, which represents the evolution of the fundamentals underpinning the risk neutral probability of default in the economy. We find strong evidence that the unprecedented monetary policy response from the Fed during the crisis period was effective in reducing market uncertainty and helped to steepen the term structure of the CDX index, thereby mitigating systemic risk concerns. The impact of stock market volatility on flattening the term premium was substantially more robust in the crisis period. We also show that equity returns make a significant contribution to the CDX term premium over the entire sample period.

Table of Contents

Chapter I Introduction.....	15
Chapter II A review of nonlinear regime switching models.....	19
II.1 Threshold Models	20
II.1.1 Basic STR Model.....	20
II.1.2 Inference in STR models	22
II.1.2.1 Test Linearity against STR	22
II.1.2.2 Specification of STR models	23
II.1.2.3 Estimation of STR models	24
II.1.2.4 Model adequacy tests of STR models.....	24
II.2 Markov-switching Models.....	24
II.2.1 Basic model.....	24
II.2.2 Estimating a basic Markov-switching model using Hamilton filter	26
II.2.3 2.2.5 Specification test of Markov-switching models	28
Bibliography	29
Chapter III A review of the models in the term structure of interest rates	32
III.1 Information contained in the term structure of interest rates.....	32
III.2 Definitions of term premia	33
III.3 Regression analysis of term premia.....	35
III.3.1 Testing on forward term premia	35
III.3.2 Testing on long-term premia	37
III.3.3 The Campbell-Shiller (1991) puzzle	37
III.3.4 Fama-Bliss (1987) and Cochrane-Piazzesi (2005) regression analysis	40
III.4 Factor models in estimating term structure of interest rates.....	50
III.4.1 Principal components	53
III.4.2 Nelson-Siegel factor model	57
III.4.2.1 Estimation of the dynamic Nelson-Siegel factor model using Kalman filter	58

III.4.2.2 Estimation results	62
III.5 No-arbitrage factor models	67
III.5.1 No-arbitrage conditions and the stochastic discount factor in term structure model	68
III.5.2 Short rate model	72
III.5.3 Multi-factor affine model	74
III.5.3.1 Discussion on the specification of market price of risk in affine term structure models.....	77
Bibliography	78
Appendix III	81
Chapter IV A Markov switching extension of the short rate model	88
IV.1 Introduction	88
IV.2 Single regime CIR model	89
IV.3 Markov switching CIR model	91
IV.3.1 Maximum Likelihood Estimation of Markov switching CIR model	92
IV.4 Empirical results.....	94
IV.5 Ex-ante pricing performance of the models	103
IV.6 Conclusion.....	105
Bibliography	106
Appendix IV	108
Chapter V Joint Modelling of the Nominal and Real Term Structure of Interest Rates and the relation between expected inflation rates and stock return risk premium.....	114
V.1 Motivation and Literature Review	115
V.1.1 Empirical behaviour of nominal term premia in no-arbitrage factor models	120
V.1.2 Empirical estimates of inflation risk premia.....	120
V.2 Inflation index-linked gilts	123
V.3 Essentially affine term structure of nominal and real interest rates.....	125
V.3.1 State equation	125
V.3.2 Nominal and real pricing kernel	126

V.3.3 Nominal bond pricing.....	127
V.3.4 Real bond pricing.....	128
V.3.5 Nominal and real term premia	129
V.3.6 Inflation risk premia and expected inflation	131
V.4 Data.....	131
V.5 Principal component analysis and simple regressions	141
V.6 Estimation	145
V.7 Empirical results	147
V.8 The relation between stock returns and expected inflation.....	160
V.8.1 The STVAR model.....	162
V.8.2 Specification tests for the STVAR models	164
V.8.2.1 Linearity test	164
V.8.2.2 Model adequacy tests.....	166
V.8.2.2.1 Test of serially correlated errors	166
V.8.2.2.2 Tests for parameter constancy	167
V.8.2.2.3 Test for no remaining nonlinearity	168
V.8.3 Empirical result.....	168
V.9 Conclusion	174
Bibliography	175
Appendix V	180
Chapter VI The stock market risk return Trade-off: evidence from six countries using centuries-long data	189
VI.1 Introduction	189
VI.2 Literature review	190
VI.3 Econometric models	193
VI.3.1 Single regime mean and variance (Model 1).....	193
VI.3.2 Single regime mean and multi-regime variance (Model 2).....	194
VI.3.3 Multi-regime mean and variance (Model 3).....	195
VI.3.4 Multi-regime mean, variance and autoregressive coefficient (Model 4).....	196

VI.3.5 Multi-regime mean, variance and autoregressive coefficient with learning (Model 5).....	196
VI.4 Data	198
VI.5 Estimation Results.....	203
VI.5.1 Australia	206
VI.5.2 Canada	213
VI.5.3 Sweden	218
VI.5.4 Switzerland.....	223
VI.5.5 UK	228
VI.5.6 US.....	232
VI.5.7 Volatility feedback effect	236
VI.5.8 The co-movements (integration) of the stock markets	238
VI.6 Conclusion.....	241
Bibliography	242
Appendix VI	245
Chapter VII A Markov Switching Unobserved Component Analysis of the CDX	
Index Term Premium	250
VII.1 Introduction.....	250
VII.2 Motivation and Methodology.....	255
VII.2.1 Stationary and Random Walk Components in State Space Representation	259
VII.2.2 State Space Model with Markov Switching Disturbances	262
VII.2.3 Estimation procedure	264
VII.3 Data	268
VII.4 Empirical Results	271
VII.4.1 Model selection tests	272
VII.4.2 Estimates of the Markov-switching unobserved component model	275
VII.4.3 VAR Analysis	279
VII.5 Conclusion	283

Bibliography	286
Appendix VII.....	289
Chapter VIII Summary and conclusion	304

List of Tables

Table III-1: UK and German Nelson-Siegel model estimates and standard errors	63
Table III-2: Descriptive statistics	81
Table III-3: Fama-Bliss regression and "complementary regression" results	82
Table III-4: Time series and cross sectional regressions of 1-year excess bond returns on all forward rates.....	83
Table III-5: Time series and cross sectional regressions of 2-year excess bond returns on all forward rates.....	84
Table IV-1: Estimated models.....	94
Table IV-2: Descriptive statistics	95
Table IV-3: Parameter estimates of various models.....	96
Table IV-4: Likelihood ratio tests	100
Table IV-5: MSE and FDT statistics	104
Table IV-6: Pesaran-Timmermann test	105
Table V-1: Descriptive statistics of nominal zero-coupon bond yields.....	136
Table V-2: Descriptive statistics of real zero-coupon bond yields.....	137
Table V-3: Principal component analysis.....	143
Table V-4: Parameter estimation results	148
Table V-5: Projection of the break-even rates on nominal yields within different sample periods	159
Table V-6: Linear approximation of LSTVAR and ESTVAR models	165
Table V-7: Nested model selection hypothesis tests	165
Table V-8: Descriptive statistics of r and $\Delta\pi$	169
Table V-9: Results for linearity test and model selection test.....	170
Table V-10: Maximum likelihood estimates of the LSTVAR models.....	171
Table V-11: Model adequacy tests	173
Table V-12: Review of recent estimations of inflation risk premia and inflation expectation	180
Table VI-1: Data sources and time span of the stock return data.....	199
Table VI-2: Descriptive statistics of the stock market return series for each country.....	199
Table VI-3: Model specification test using Hansen 1992 test procedure.....	205
Table VI-4: Estimation results (Australia)	206
Table VI-5: Ljung-Box Q statistics for autocorrelation effect and Q^2 statistics for ARCH effect on the residuals (Australia).....	207
Table VI-6: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - Australia	209
Table VI-7: Estimation result (Canada).....	214
Table VI-8: Ljung-Box Q statistics for autocorrelation effect and Q^2 statistics for ARCH effect on the residuals (Canada)	215
Table VI-9: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - Canada.....	215
Table VI-10: Estimation result (Sweden).....	218
Table VI-11: Ljung-Box Q statistics for autocorrelation effect and Q^2 statistics for ARCH effect in the residuals (Sweden)	220
Table VI-12: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - Sweden	221

Table VI-13: Estimation result (Switzerland)	223
Table VI-14: Ljung-Box Q statistics for autocorrelation effect and Q^2 statistics for ARCH effect in the residuals (Switzerland)	224
Table VI-15: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - Switzerland.....	225
Table VI-16: Estimation result (UK).....	228
Table VI-17: Ljung-Box Q statistics for autocorrelation effect and Q^2 statistics for ARCH effect on the residuals (UK)	229
Table VI-18: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - UK.....	230
Table VI-19: Estimation result (US)	232
Table VI-20: Ljung-Box Q statistics for autocorrelation effect and Q^2 statistics for ARCH effect on the residuals (US)	233
Table VI-21: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - US.....	235
Table VI-22: Coherence in regimes between pairs of countries.....	239
Table VI-23: Estimation results of Model 5 for each country	245
Table VII-1: Mean and standard deviation of the CDX 5-year (CDX5Y), CDX 10-year (CDX10Y) and Term Premium (TP) for different sample periods	258
Table VII-2: Descriptive statistics of the data series	270
Table VII-3: Model selection results	273
Table VII-4: Likelihood ratio tests on constraint $\rho_{12} = \rho_{21} = 0$	274
Table VII-5: Likelihood ratio tests within groups	274
Table VII-6: Residual diagnostic tests.....	274
Table VII-7: Estimation results of Model 8.....	276
Table VII-8: Generalized variance decomposition of DIFRW and STAT in the pre-financial crisis period(September 13, 2004 to January 03, 2008)	282
Table VII-9: Generalized variance decomposition of DIFRW and STAT in the post-financial crisis period (January 04, 2008 to July 23, 2009).....	283

List of Figures

Figure III-1: Unrestricted and restricted coefficients	44
Figure III-2: Forecast and realized average excess bond return in the UK	45
Figure III-3: Forecast and realized average excess bond return in the US	45
Figure III-4: Forecast and realized average excess bond return in German	46
Figure III-5: UK 2-year holding period regression coefficients.....	48
Figure III-6: US 2-year holding period regression coefficients.....	48
Figure III-7: German 2-year holding period regression coefficients.....	49
Figure III-8: US 1 to 5 years Zero-Coupon-Yield level and slope (06/1952–12/2008)	51
Figure III-9: UK 1 to 10 years Zero-Coupon-Yield level and slope (01/1970–09/2009).....	51
Figure III-10: German 1 to 10 years Zero-Coupon-Yield level and slope (09/1972–09/ 2009)	52
Figure III-11: Replicate the US zero coupon bond yields using principal components	55
Figure III-12: US replication residuals square	55
Figure III-13: Replicate the UK zero coupon bond yields using principal components	56
Figure III-14: UK replication residuals square.....	56
Figure III-15: Replicate the German zero coupon bond yields using principal components ...	57
Figure III-16: German replication residuals square.....	57
Figure III-17: What do Nelson-Siegel factor loadings look like	58
Figure III-18: UK Nelson-Siegel factors	64
Figure III-19: German Nelson-Siegel factors	65
Figure III-20: Residual square of the fitted UK yields.....	65
Figure III-21: Residual square of the fitted German yields.....	66
Figure III-22: Observed and fitted UK yields	66
Figure III-23: Observed and fitted German yields	67
Figure IV-1: UK zero-coupon bond yields (% , January 1970 to September 2010)	97
Figure IV-2: Smoothed state probabilities of regime 0 for various models	98
Figure IV-3: Observed 3-month rates and fitted 3-month rates from models	99
Figure IV-4: Observed 5-year rates and fitted 5-year rates from models.....	99
Figure IV-5: Observed 10-year rates and fitted 10-year rates from models.....	100
Figure IV-6: Snap shot of the yields between January 2003 and September 2010	101
Figure IV-7: Recursive AIC and BIC.....	102
Figure V-1: UK nominal zero-coupon yields for various maturities.....	133
Figure V-2: UK RPI(Percentage change over 12 months - all items)	134
Figure V-3: UK zero-coupon real yields for various maturities.....	135
Figure V-4: UK average nominal yield curve with +/- 2 S.E. bounds	138
Figure V-5: UK nominal yield volatility curve with +/- 2 S.E. bounds	138
Figure V-6: UK average real yield curve with +/- 2 S.E. bounds	139
Figure V-7: UK real yield volatility curve with +/- 2 S.E. bounds	139
Figure V-8: UK nominal yield autocorrelation curve with +/- 2 S.E. bounds.....	140
Figure V-9: UK real yield autocorrelation curve with +/- 2 S.E. bounds.....	141
Figure V-10: Factor loadings of nominal yields on principal components	143
Figure V-11: Factor loadings of real yields on principal components	144
Figure V-12: Filtered three unobserved factors.....	149
Figure V-13: Observed and fitted nominal yields (3-month, 1-year, 3-year and 5-year).....	149
Figure V-14: Observed and fitted nominal yields (6-year, 7-year, 8-year and 10-year)	150

Figure V-15: Observed and fitted real yields (4-year, 5-year, 6-year and 7-year)	150
Figure V-16: Observed and fitted real yields (8-year, 9-year and 10-year)	151
Figure V-17: 5-year nominal yields and term premia	152
Figure V-18: 10-year nominal yields and term premia	152
Figure V-19: 5-year real yields and term premia	153
Figure V-20: 10-year real yields and term premia	153
Figure V-21: 10-year break-even rate and inflation risk premia	154
Figure V-22: Term structure of the average nominal and real term premia with +/- 2 S.E. bounds	155
Figure V-23: Term structure of the volatility of nominal and real term premia with +/- 2 S.E. bounds	155
Figure V-24: Term structure of the autocorrelation of nominal and real term premia with +/- 2 S.E. bounds.....	156
Figure V-25: Term structure of the inflation risk premia (average, volatility and autocorrelation curves) with +/- 2 S.E. bounds	156
Figure V-26: The observed and expected 1 month inflation rates.....	157
Figure V-27: Observed and expected annual inflation rates	158
Figure V-28: 4-year to 10-year expected inflation rate	159
Figure V-29: Monthly stock return risk premium (FTSE100) and changes of monthly expected inflation rates.....	169
Figure V-30: Logistic transition function from LSTVAR model with $s_t = \Delta\pi_{t-4}$	172
Figure V-31: Logistic transition function from LSTVAR model with $s_t = r_{t-5}$	173
Figure VI-1: Australian stock market index and stock return	200
Figure VI-2: Canadian stock market index and stock return	200
Figure VI-3: Swedish stock market index and stock return	201
Figure VI-4: Switzerland stock market index and stock return.....	201
Figure VI-5: The UK stock market index and stock return	202
Figure VI-6: The UK stock market index and stock return (Sep. 1902 to Sep. 2007)	202
Figure VI-7: The US stock market index and stock return.....	203
Figure VI-8: The US stock market index and stock return (Sep. 1902 to Sep. 2007).....	203
Figure VI-9: Australian stock market return series and Model 2-AR(1)'s smoothed posterior probability of observing a high volatility regime (2-regime)	209
Figure VI-10: Australian stock market return series and Model 2-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)	210
Figure VI-11: Australian stock market return series and Model 2-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)	211
Figure VI-12: Australian stock market return series and Model 2-AR(0)'s smoothed posterior probability of observing a high volatility regime (3-regime)	212
Figure VI-13: Australian stock market return series and Model 3-AR(0)'s smoothed posterior probability of observing a high volatility regime (3-regime)	213
Figure VI-14: Canadian stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a high volatility regime (2-regime)	215
Figure VI-15: Canadian stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)	217
Figure VI-16: Canadian stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)	218

Figure VI-17: Sweden stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)	222
Figure VI-18: Sweden stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)	222
Figure VI-19: Switzerland stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a high volatility regime (2-regime)	225
Figure VI-20: Switzerland stock market return series and Model 4-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)	227
Figure VI-21: Switzerland stock market return series and Model 4-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)	227
Figure VI-22: The UK stock market return series and Model 4-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)	231
Figure VI-23: The UK stock market return series and Model 4-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)	232
Figure VI-24: The US stock market return series and Model 2-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)	235
Figure VI-25: The US stock market return series and Model 2-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)	236
Figure VI-26: Time evolution of the coherence in regimes between pairs of countries (1)...	240
Figure VI-27: Time evolution of the coherence in regimes between pairs of countries (2)...	241
Figure VI-28: Time evolution of the coherence in regimes between pairs of countries (3)...	241
Figure VII-1: CDX-5 year, CDX-10 Year and CDX term premium.....	270
Figure VII-2: Actual plots of: <i>FFR</i> , <i>SLOPE101</i> , <i>SP500RTN</i> and <i>VIX</i>	271
Figure VII-3: Snapshot of the stationary component in its low volatility regime	277
Figure VII-4: Stationary component and filtered state probability of its high volatility regime	289
Figure VII-5: Random walk component and filtered state probability of its high volatility regime	290
Figure VII-6: Conditional variance of the term premium and filtered state probability of high volatility regime for each component.....	291
Figure VII-7: Accumulative generalized impulse response functions of the VAR model in the pre-financial crisis period (September 13, 2004 to January 03, 2008).....	292
Figure VII-8: Accumulative generalized impulse response functions of the VAR model in the post-financial crisis period (January 04, 2008 to July 23, 2009).....	293

Chapter I Introduction

In recent years a major class of nonlinear time series models - regime switching model - has become a popular workhorse in many economic and financial studies. For example, studies dedicated to business cycles, equity return forecasting, term structure of interest rates and exchange rates volatility modelling, have been adopting increasingly a regime switching approach to describe the time series properties of economic and financial variables. Why the regime switching models are so attractive to researchers? Arising from practical aspects of modelling dynamic economic and financial time series, researches often observe variables undergo either/both permanent structural changes or/and recurrent temporary changes over a long sample period. The most well-known example of this type is the repeatedly appearance of expansion and recession phrases in business cycles. The apparent time inconsistency patterns make researchers believe that the constant parameter time series models may not be able to incorporate the observed regime changes in practice. One way to solve this problem is to build a stochastic process into the time series model that is capable to describe the regime switches in the mean or/and variance of the variables being studied. Although we do not directly observe these regime switches at any point in time, we may draw inferences on the regimes in operation of a known stochastic process, and hence the possible future switches in forecast.

This dissertation comprises a collection of four empirical essays in nonlinear time series analysis of economic/financial variables. Two of them involve the term structure of interest rates modelling. The term structure of interest rates plays a central role in economy, both theoretically and practically. While policy makers conduct monetary policy by shifting interest rates at the short end of yield curve, longer term yields give markets' reflection of future economic health. When expectations about future rates change, the yield curve reacts as bonds reprice to reflect these expectation changes. Moreover, longer term yields also compound liquidity risk premium and nonlinear convexity premium. Therefore, the movements in longer end of the term structure reflect not only the revision of expectation but also various risk premiums.

Giving the importance of interest rate modelling, the related econometric estimation and forecasting methods have received tremendous attention in recent years. One popular

approach is to incorporate a regime switching feature into the term structure models, which allows the pricing kernel to be regime dependent. Many studies in interest rate modelling have shown empirical evidence of regime switches in short rates. However, none of them (except Driffill, *et al.*(2009)) considers the impact of different model specifications on the performance of model's forecasting ability. In Chapter IV, we estimate a nonlinear Markov switching CIR model on the term structure of UK interest rates and evaluate different parameterizations of the model in terms of real-time one-step-ahead bond pricing performance. We conduct a series of model specification tests on a battery of models that employ different parameter restrictions on the short rate equation. By doing so, we provide a comprehensive Markov switching CIR model specification analysis on the UK data that varies from January 1970 to September 2010.

The issue of the term structure modelling is further discussed in Chapter V, where we jointly model the nominal and real term structure of UK interest rates. The expected rate of inflation is an important economic variable to both policy makers and investors. On the one hand, it measures the credibility of central bank's monetary policy aiming at anchoring the long term inflation rate around its target level. On the other hand, the inflation risk premia demanded by investors, if non-negligible, would contaminate the interpretation of the break-even inflation rate as a measure of the credibility widely used. However, due to the unobserved nature of real rates, expected inflation rates and the inflation risk premia, making correct identification of these quantities in reality is difficult. To overcome this difficulty, we decompose the nominal term structure of interest rates into real interest rates, expected inflation rates and inflation risk premia by employing a three-factor essentially affine no-arbitrage term structure model on the UK data. The model implied expected inflation rate is then used in the subsequent nonlinear analysis of the relation between stock return premiums and inflation rates.

In Chapter VI, we investigate the risk-return trade-off in stock market. In this study, we provide additional insight into the nature of the aggregate stock market volatilities and its relations to expected returns, in a Markov switching model using centuries-long aggregate stock market data from six countries (Australia, Canada, Sweden, Switzerland, the United Kingdom and the United States). We consider a wide range of models in this study. By relaxing restrictions on the multi-regime switching parameters in each step, we estimated a series of models either allowing each of the expected mean and volatility of the stock return

to switch or permitting both moments of the stock return to shift regimes over time. In contrast to most studies in Markov switching modelling of stock returns that assuming two regimes, we also consider three-regime specification for the variation of stock returns. This seemingly subtle complication, as compared to the conventional two regimes assumption, allows a richer inter-temporal relation between the expected mean and volatility of the stock return. We also investigate into the "volatility feedback effect" hypothesis that has been documented extensively in literature. We make different assumptions about the information available to market investors. By allowing investors to learn about the current prevailing volatility regime based on different information set prior to the current trading period, we are able to see the effect on the additional risk premium required by investors in anticipating future volatility changes.

In the final chapter, we estimate a Markov switching unobserved component model to explain the evolution of the term premium for the North American CDX index. The term premium of the CDX index, which is measured as the difference between a longer maturity CDX index series and a shorter maturity one (e.g. the difference between the CDX 10-Year index and the CDX 5-Year index), can be viewed as representing the 5-year forward uncertainty regarding corporate default after the next 5 years. Accordingly, the CDX term premium can be interpreted as an early warning market indicator of improvement or deterioration in macroeconomic conditions 5 years hence. Our interest in this study is to study how the factors themselves (not the factor loadings) drive the dynamics of the term premium. To characterize the observed patterns of volatility jumps on the CDX index term premium, we allow on the innovation terms a regime switching process, following two distinct first-order Markov chain variables. We consider an appropriately specified Markov Switching Unobserved Components model as a reliable measure of volatility dynamics of the CDX index spread curve and investigate the presence and significance of both monetary policy adjustments and stock market returns for the US economy over the sample period September 2004 - July2009.

This dissertation is organized as follows. In Chapter II, we give a short review of the nonlinear models used in this dissertation. Chapter III reviews the models of the term structure of interest rates in a linear modelling fashion. Chapter IV presents the study of the Markov switching extension of the short rate model on the UK data. Chapter V discusses the joint modelling of the UK's nominal and real term structure of interest rate with a subsequent

analysis on the relation between stock return premium and the expected inflation rates based on a nonlinear smooth transition vector autoregressive model. Chapter VI investigates the risk-return trade-off in stock market in a flexible Markov switching modelling framework by using century-long aggregate stock market data from six countries. Chapter VII presents the Markov switching unobserved component modelling of the CDX term premium in the US market. Finally, Chapter VIII concludes.

Chapter II A review of nonlinear regime switching models

Generally speaking, regime switching models can be classified into two groups: threshold regime switching models, and Markov regime switching models. The main difference between the two groups of models is that they adopt different assumptions on the stochastic process of the regime switching variable. Tong (1983) proposed a threshold regime switching model that assumes the regime switching variable is an unobserved threshold, and regimes are defined as above or below the value of the threshold variable. Later, developments in Chan and Tong (1986) and Tong (1990) augmented the threshold variable to be a lagged endogenous variable, which constitutes the so-called threshold autoregressive (TAR) model. By connecting the TAR model with the “smooth transition” idea in Bacon and Watts (1971), researchers devised a general smooth transition regression (STR) model. The empirical appeal of this STR model is that it allows economic variables to transit gradually from one regime to another rather than suddenly change at a fixed threshold value. Maddala (1977) replaced the smoothing function from cumulative distribution function to logistic function, which became a popular choice for the transition function thereafter. With the combination of the STR model and the assumption that the threshold variable is a lagged endogenous variable, Terasvirta (1994) came up with the popularized smooth autoregressive (STAR) model.

Markov regime switching models, which were initially introduced by Goldfeld and Quandt (1973) and popularized by Neftçi (1982) and Hamilton (1989b), assume that the switching variable follows a Markov chain process. In his empirical study of the U.S. business cycles, Hamilton (1989b) assumes fundamental macroeconomic variables, such as GNP and real output growth, behave differently depending on which state the economy is in at the time. The shifts of states between expansion and contraction of the economy are governed by an unobservable regime switching variable that follows a Markov chain process. The regime switching variable takes two values, 0 and 1, to represent contraction and expansion states of the economy, respectively. Markov regime switching models were firstly invoked in macroeconomic time series studies, particularly in business cycle literatures. Since then, a number of empirical studies have established its statistical advances over the traditional constant parameter time series models. Recent developments in the Markov

regime switching literature on business cycle modelling are surveyed in Hamilton and Raj (2002) and Raj (2002).

II.1 Threshold Models

Early development of the threshold model is based on the discrete threshold transition model. Bacon and Watts (1971) first suggested the phrase “smooth transition”, and showed how the local linear equations can be modelled as a smooth change from one linear extreme to the other when linked to a continuous transition variable. In econometric literature, Goldfeld and Quandt (1973) is the first to propose the following model

$$\begin{aligned}
 y_t &= \alpha_1 [1 - D(z_t)] + \alpha_2 D(z_t) \\
 &+ \{ \beta_1 [1 - D(z_t)] + \beta_2 D(z_t) \} x_t + [1 - D(z_t)] u_{1t} + D(z_t) u_{2t} \\
 D(z_t) &= \begin{cases} 1, & z_t \geq c \\ 0, & z_t < c \end{cases}
 \end{aligned} \tag{II.1}$$

It says when the transition variable $z_t \geq c$, the regression intercept will be α_2 , slope parameter will be β_2 , and the error term will be u_{2t} , otherwise the regression is run on the set of parameters with subscript of 1. In terms of estimation, they suggest a simplified approximation of $D(z_t)$, that is a cumulative distribution function with mean c and variance σ_2 , as depicted by (II.2):

$$D(z_t) = -\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{z_t} \exp\left(-\frac{1}{2} \frac{(z-c)^2}{\sigma^2}\right) dz. \tag{II.2}$$

II.1.1 Basic STR Model

The Basic STR model takes the following specification

$$\begin{aligned}
 y_t &= x_t' \alpha + (x_t' \beta) G(\gamma, c; s_t) + u_t, \\
 t &= 1, 2, \dots, T,
 \end{aligned} \tag{II.3}$$

where x_t is a vector of explanatory variables consisting of lagged endogenous and exogenous variables, i.e. $x_t = (1, x_{1t}, \dots, x_{pt})' = (1, y_{t-1}, \dots, y_{t-m}; z_{1t}, \dots, z_{mt})'$; α and β are two different sets of parameters, i.e. $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_p)'$ and $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$; u_t is a sequence of i.i.d. errors; G is a continuous transition function customarily bounded between 0 and 1.

Conventionally, G is a function of three elements, among which γ is a positive identification restriction that indicates the speed of transition from one regime to the other, c is a location parameter that determines where the transition takes place, and s_t is the transition variable. There are many ways to specify the transition variable s_t . Terasvirta (1994) defines the STAR model by specifying the transition variable as a lagged endogenous variable, which is $s_t = y_{t-d}$ with $d > 0$. Alternatively, the transition variable can be an exogenous variable, like $s_t = z_t$, or a nonlinear function of the lagged endogenous variable, like $s_t = h(\tilde{x}_t; \kappa)$ with $\tilde{x}_t = (y_{t-1}, \dots, y_{t-m})$ for some nonlinear function h conditioned on the parameter vector κ . Furthermore, the transition variable can be a function of time trend, i.e. $s_t = t$, as in Lin and Terasvirta (1994) that allows smooth time varying parameters.

From (II.3) we see the STR model nests two linear equations, which represent the two local regimes: $y_t = x_t' \alpha + u_t$ when $G = 0$, and $y_t = x_t' (\alpha + \beta) + u_t$ when $G = 1$. If $0 < G < 1$, the model describes the continuum of states in between the two local regimes. A practical issue about the STR models is how we define the transition function G . Here we list few specification of this transition function following Terasvirta (1998).

$$\text{LSTR1:} \quad G_1(\gamma, c; s_t) = \frac{1}{1 + \exp[-\gamma(s_t - c)]} \quad (\text{II.4})$$

Equation (II.4) is called the logistic STR (LSTR1) model. This transition function is monotonically increasing with the transition variable s_t . The speed of transition as represented by γ is always positive. If $\gamma \rightarrow \infty$, it implies an immediate switch of regimes as in the discrete TAR and SETAR models. If $\gamma \rightarrow 0$, the transition function approaches the constant 0.5, and hence it reduces to a linear model of $y_t = x_t' \left(\frac{\alpha + \beta}{2} \right) + u_t$.

$$\text{LSTR2:} \quad G_2(\gamma, c; s_t) = \frac{1}{1 + \exp[-\gamma(s_t - c_1)(s_t - c_2)]} \quad (\text{II.5})$$

$$c_1 \leq c_2$$

Equation (II.5), which we shall call it quadratic logistic function(LSTR2), is an alternative specification of the transition function proposed by Eitrheim and Terasvirta

(1996a). By calling "quadratic", it is easy to see that G_2 is symmetric about the location $\frac{c_1 + c_2}{2}$.

$$\text{ESTR:} \quad G_3(\gamma, c; s_t) = 1 - \exp\left[-\gamma(s_t - c)^2\right] \quad (\text{II.6})$$

Equation (II.6), the exponential STR (ESTR) model, describes another popular specification of the transition function, which allows re-switching but with a more rapid speed. This transition function, which is bounded between $\lim_{s_t \rightarrow \pm\infty} G_3 = 1$ and 0, is symmetric at the transition point. Comparing ESTR with LSTR2, these two transition functions have similar shapes when γ is small and $c_2 - c_1$ is close to zero.

II.1.2 Inference in STR models

II.1.2.1 Test Linearity against STR

We start with redefining the aforementioned transition functions as

$$\begin{aligned} y_t &= x_t' \alpha^* + (x_t' \beta^*) G_i^*(\gamma, c; s_t) + u_t \\ G_i^*(\gamma, c; s_t) &= G_i(\gamma, c; s_t) - 0.5, \text{ for } i = 1, 2 \end{aligned} \quad (\text{II.7})$$

The null hypothesis of linearity requires $G_i^* = 0$, which amounts to $\gamma = 0$ or $\beta^* = 0$. However, as described in Luukkonen, Saikkonen and Terasvirta (1988), parameters c and β are not identified under the null hypothesis which, consequently, leads to inconsistent estimation of these two parameters. Furthermore, the statistics for likelihood ratio test, Lagrange multiplier test and Wald test do not have their correct standard asymptotic distributions under the null. To deal with this problem, we need to write the redefined transition functions in their appropriate Taylor approximations evaluated at $\gamma = 0$. Following Terasvirta (1998), we use the LSTR1 model as an illustration and write

$$\begin{aligned} TA_1 &= \varphi_0 + \varphi_1 s_t + R_1(\gamma, c; s_t), \\ y_t &= x_t' \alpha^{**} + (x_t s_t)' \beta^{**} + u_t^*, \\ u_t^* &= u_t + (x_t' \beta^*) R_1(\gamma, c; s_t), \end{aligned} \quad (\text{II.8})$$

where φ_0 and φ_1 are the coefficients of the first two terms of the Taylor expansion, and $R_1(\gamma, c, s_t)$ is the remainder. Now the null hypothesis becomes $H_0 : \beta^{**} = 0$ against the alternative $H_1 : \beta^{**} \neq 0$, which can be tested straight forwardly in a Lagrange multiplier test.

Another point to note is the problem of low power in the LM test, as described in Terasvirta (1998). A general solution to this problem requires the augmentation of a further order from the Taylor expansion series to the reconstructed model, which is

$$y_t = x_t' \alpha^{**} + (x_t' s_t) \beta^{**} + (x_t' s_t^2) \xi^{**} + (x_t' s_t^3) \zeta^{**} + u_t^*. \quad (\text{II.9})$$

Consequently, under the null hypothesis of linearity, a joint F-test is performed against the alternative.

II.1.2.2 Specification of STR models

In a way similar in spirit to Box and Jenkins (1970), Granger and Terasvirta (1993)¹ suggest a modelling cycle for STR models. This modelling cycle includes finding a suitable transition variable, estimating procedures, and adequacy testing. The procedure is as follows: first, we begin with the conventional linear modelling using Box-Jenkins's method to see if a linear model is adequate or not; second, if the linear model is not a good approximation of the data, we try the linearity test against STR models with each of the potential transition variables tested; third, if more than one potential transition variables are qualified, we tend to select the most statistically significant one as the transition variable unless the prior information from suggested economic theory is known; fourth, after the work of selecting a suitable transition variable is done, we need to choose the type of the model to estimate, that is defining a sequence of null hypothesis tests based on (II.9) and carrying out the corresponding decision rules as

$$\begin{aligned} H_{04} : \zeta^{**} &= 0 \\ H_{03} : \xi^{**} &= 0 \mid \zeta^{**} = 0 \\ H_{02} : \beta^{**} &= 0 \mid \xi^{**} = \zeta^{**} = 0. \end{aligned} \quad (\text{II.10})$$

All hypotheses are tested under the framework of F-test, if the most significant rejection happens on H_{03} , we choose LSTR2 and carry on with the additional test on whether $c_1 = c_2$

¹ See Chapter 7 for details.

to determine if ESTR is better than LSTR2; if the most significant rejection happens on either H_{04} or H_{02} , we choose LSTR1².

II.1.2.3 Estimation of STR models

STR models can be estimated using nonlinear least squares or maximum likelihood methods assuming normally distributed errors. Associated with the maximum likelihood method, nonlinear optimization algorithms are used, such as the Newton-Raphson algorithm and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Topics on numerical optimisation for maximum likelihood methods can be found in Hendry (1995), Judge (1985) and Quandt (1984). Additional remark need to be made on the numerical methods is that the speed of transition parameter γ is not scale free. To eliminate this magnitude dependence, we need to divide the exponent of the transition function by the sample standard deviation of s_t in LSTR1 mode and sample variance of that in LSTR2 model.

II.1.2.4 Model adequacy tests of STR models

Once the STR models have been estimated, we need to consider the specification tests, which check the validity of the underlying assumptions in models. Specification tests for STR models are in common with the test against linearity that they are both designed against parametric alternatives. Eitrheim and Terasvirta (1996a) and Terasvirta (1998) have considered the specification tests in ST(A)R models. Generalizations of those tests in a LM framework are straight forward, and consist of three types of tests - no-error autocorrelation test, no-additive nonlinearity test, and parameter constancy test. Lutkepohl and Kratzig (2004)³ have given a detailed account of the three tests. The related asymptotic normality and consistency condition from maximum likelihood estimation can be find in Wooldridge (1994) and Escribano and Mira (1995).

II.2 Markov-switching Models

II.2.1 Basic model

Let's start with the simplest model of describing an economic variable evolves over time - the first-order autoregressive model $y_t = \alpha_1 + \beta_1 y_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$. Let's assume this is the dynamic under the economic booming condition, whereas when the economic

² For alternative tests based on choosing-logics, Terasvirta (1998) gives the details. For detailed procedures on the above tests, please refer to Granger and Terasvirta (1993) and Terasvirta (1994). Also one can find the simulation results of the property of using the above choosing-logics in Terasvirta (1998).

³ See Chapter 6 for details.

recession comes, the previous parameters could not sustain and alternatively we may demand another set of parameters to present the new behaviour of this economic variable as $y_t = \alpha_2 + \beta_2 y_{t-1} + \varepsilon_t$. If one would like to write the two equations into a compact one, we have

$$y_t = \alpha_{s_t} + \beta_{s_t} y_{t-1} + \varepsilon_t, \quad (\text{II.11})$$

where s_t is the regime indicator that takes value 1 when the economy is booming and equals 2 when in recession. A probabilistic view of the regime changes would be obtained by defining the following probabilistic Markov chain

$$p_{ij} = \Pr(s_t = j \mid s_{t-1} = i, s_{t-2} = l, \dots, y_{t-1}, y_{t-2}, \dots) = \Pr(s_t = j \mid s_{t-1} = i). \quad (\text{II.12})$$

Equations (II.11) and (II.12) together constitute the simplest version of Markov-switching model. In such a model, we observe y_t and make inference about s_t on observing new y_t , that is y_{t+1} , and so on.

Equation (II.12) specifies a complete probability distribution for the switching variable. The majority of the work on Markov-switching models has assumed the transition probabilities to be independent from the lagged endogenous variable, and this is why the Markov-switching models are often called “exogenous” switching models compared to “endogenous” threshold models. In general, this Markov chain process allows regimes to be visited randomly, that is in any order and could be visited more than once. However, as Chib (1998) noted, restrictions can be placed on the transition probability in a way that the model becomes a structural break model, so that each regime can be visited only once and each transition becomes a changing point.

To estimate a Markov-switching autoregressive model, one needs to make a further assumption on the structure of the transition probabilities that govern the switching variable. As the transition probabilities have to be constrained between 0 and 1, a logistic function is usually specified for s_t . For example, in a two-regime model (1 represents boom and 2 represents recession),

$$\begin{aligned}
p_{11} &= \Pr(s_t = 1 | s_{t-1} = 1) = \frac{\exp(p_0)}{1 + \exp(p_0)} \\
p_{12} &= 1 - p_{11} \\
p_{22} &= \Pr(s_t = 2 | s_{t-1} = 2) = \frac{\exp(q_0)}{1 + \exp(q_0)} \\
p_{21} &= 1 - p_{22},
\end{aligned}$$

where p_0 and q_0 are some unconstrained parameters.

To estimate a Markov-switching model, we must achieve two tasks. First, we estimate the parameters by maximising the likelihood function. Second, we draw inference on s_t given the maximized parameters. In order to construct the likelihood function, we need to employ the Hamilton filter, which is a recursive algorithm to compute the conditional densities.

II.2.2 Estimating a basic Markov-switching model using Hamilton filter

We first identify the number of parameters we need to estimate in (II.11) and (II.12). If we assume there are two regimes, that is $s_t = 1$ and $s_t = 2$ representing regime 1 and regime 2, respectively, we will have a parameter vector $\theta = (\alpha_1, \beta_1, \alpha_2, \beta_2, \sigma, \gamma_1, \gamma_2)$ that need to be estimated. Then for a given value of θ , the conditional log likelihood function is given by

$$L(\theta) = \sum_{t=1}^T \log f(y_t | \Omega_{t-1}; \theta). \quad (\text{II.13})$$

We capture the conditional densities $f(y_t | \Omega_{t-1}; \theta)$ at each $t = 1, 2, \dots, T$, recursively, by using the Hamilton filter, which begins with the known initial⁴ value of $\Pr(s_{t-1} = i | \Omega_{t-1}; \theta)$ ⁵, then we construct

$$\Pr(s_t = j | \Omega_{t-1}; \theta) = \sum_{i=1}^2 \Pr(s_t = j | s_{t-1} = i, \Omega_{t-1}; \theta) \times \Pr(s_{t-1} = i | \Omega_{t-1}; \theta) \quad (\text{II.14})$$

and

⁴ Here the initial time is when $t-1=0$.

⁵ This is the posterior probability that $s_{t-1} = i$ conditioned on all information observed up till time $t-1$.

$$f(y_t | \Omega_{t-1}; \theta) = \sum_{j=1}^2 f(y_t | s_t = j, \Omega_{t-1}; \theta) \times \Pr(s_t = j | \Omega_{t-1}; \theta). \quad (\text{II.15})$$

From (II.15), the first term in the summation denotes the conditional density of y_t given $s_t = j$, the second term in the summation is like a weighting function used to determine how much weight should be given to $f(y_t | \Omega_{t-1}; \theta)$ from each regime. By assuming normal distribution for ε_t in each regime, we have

$$f(y_t | s_t = j, \Omega_{t-1}; \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y_t - \alpha_j - \beta_j y_{t-1})^2}{2\sigma^2}\right]. \quad (\text{II.16})$$

Substitute (II.16) into (II.15), we have $f(y_t | \Omega_{t-1}; \theta)$, and the next step is to update (II.14) and (II.15) to calculate $f(y_{t+1} | \Omega_t; \theta)$. Meanwhile we need to update $\Pr(s_t = j | \Omega_{t-1}; \theta)$ on observing y_t in order to get $\Pr(s_t = j | \Omega_t; \theta)$, this can be done via the Bayes' rule as

$$\Pr(s_t = j | \Omega_t; \theta) = \frac{f(y_t | s_t = j, \Omega_{t-1}; \theta) \times \Pr(s_t = j | \Omega_{t-1}; \theta)}{f(y_t | \Omega_{t-1}; \theta)}. \quad (\text{II.17})$$

Therefore, the Hamilton filter iteratively calculates (II.14) to (II.17) for $t=1, 2, \dots, T$ and finally gives the log likelihood function of (II.13). Maximum likelihood estimates of $\hat{\theta}_{MLE}$ then will be found by convenient numerical optimisation procedure.

With the Hamilton filter procedure, the estimation of Markov-switching model becomes relatively easier to implement. The only element left unsolved is the initial values of the transition probabilities needed to start the filter. The usual practice, which is discussed in Hamilton (1989b), is to assume the Markov process of the switching variable is an ergodic Markov chain. Therefore, the exact evaluation of the probability does not need to be involved, but simply to set them equal to the unconditional probabilities, that is $\Pr(s_{t-1} = j | \Omega_{t-1}; \theta) = \Pr(s_0 = j | \Omega_0; \theta) = \Pr(s_0 = j)$. Taking the two-regime case as an example, the unconditional probabilities are

$$\begin{aligned}\Pr(s_0 = 1) &= \frac{1 - p_{22}}{2 - p_{11} - p_{22}} \\ \Pr(s_0 = 2) &= \frac{1 - p_{11}}{2 - p_{11} - p_{22}}\end{aligned}\tag{II.18}$$

Alternatively, as explained in Hamilton (1994a) and Kim and Nelson (1999), the initial values can be treated as additional parameters to be estimated.

According to the Hamilton filter, the conditional probabilities $\Pr(s_t = j | \Omega_t; \theta)$ are the by-products of maximising the likelihood functions, which are often called “filtered probabilities”. Inference of the switching variable can be drawn on those filtered probabilities conditioning on $\theta = \hat{\theta}_{MLE}$. Furthermore, the inference can be based on the so-called “smoothed probabilities” (Kim (1994) and Kim and Nelson (1999)), which are computed using all available information (the whole sample).

II.2.3 Specification test of Markov-switching models

So far, we have assumed the number of regimes is two, mostly due to the practical reasons that we often observe time series variables switch between two distinct states, for example, booms and recessions in business cycle, high and low volatilities in financial securities. However, the real number of regimes is not known in general. To test how many regimes exist in a time series, let's consider the original basic model of (II.11). We want to test the null hypothesis of only one regime against the alternative hypothesis of two regimes, the likelihood ratio test statistics can be used, that is

$$LR = 2 \left[L(\hat{\theta}_{MLE}^2) - L(\hat{\theta}_{MLE}^1) \right] \sim \chi_4^2,$$

where $\hat{\theta}_{MLE}^2$ and $\hat{\theta}_{MLE}^1$ are the 2-regime's and 1-regime's maximum likelihood estimators, respectively. As the number difference of parameters under null and alternative is 4 in this case, the LR statistics will be distributed to χ_4^2 . However, the problem of unidentified parameters under null hypothesis would invalidate the test statistics. As explained in Davies (1977), the standard conditions, which all parameters need to be identified under null hypothesis, to create asymptotic Chi-square distribution may lead misleading results. To justify the tests asymptotically under null, the Hansen (1992) and Garcia (1998)'s techniques of computing valid LR statistics can be adopted. Recently, Carrasco, Hu and Ploberger (2004)

developed a general parameter constancy test against alternative Markov-switching models. As in the test no requirement of the estimation under alternative hypothesis is needed, which brings simplicity. If we want to use Bayesian approach, the relevant Bayes Factors and posterior odds ratios can be computed to compare different models with different regime specifications. Examples use these comparison metrics are Chib (1995) and Kim and Nelson (2001).

Bibliography

- Bacon, David W., and Donald G. Watts, 1971, Estimating the transition between two intersecting straight lines, *Biometrika : a journal for the statistical study of biological problems* 58, 525-534.
- Box, George E. P., and Gwilym Meirion Jenkins, 1970. *Time series analysis : Forecasting and control* (Holden-Day ; [Maidenhead] : Distributed by McGraw-Hill, San Francisco ; London).
- Carrasco, M., L. Hu, and W. Ploberger, 2004, Optimal test for markov switching, *Working paper, University of Rochester*.
- Chan, K. S., and H. Tong, 1986, On estimating thresholds in autoregressive models, *Journal of Time Series Analysis* 7, 179-190.
- Chib, Siddhartha, 1995, Marginal likelihood from the gibbs output, *Journal of the American Statistical Association* 90, 1313-1321.
- Chib, Siddhartha, 1998, Estimation and comparison of multiple change-point models, *Journal of Econometrics* 86, 221-241.
- Davies, R. B., 1977, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* 64, 247-254.
- Eitrheim, Oyvind, and Timo Terasvirta, 1996a, Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* 74, 59-76.
- Escribano, A., and S. Mira, 1995, Nonlinear time series models: Consistency and asymptotic normality of nls under new conditions, *Universidad Carlos III de Madrid, Statistics and Econometrics Series 14, Working Paper* 95-42.
- Garcia, R., 1998, Asymptotic null distribution of the likelihood ratio test in markov switching models, *International Economic Review* 39, 763-788.

- Goldfeld, S. M., and R. E. Quandt, 1973, A markov model for switching regressions, *Journal of Econometrics* 1, 3-15.
- Granger, C. W. J., and Timo Terasvirta, 1993. *Modelling nonlinear economic relationships* (Oxford University Press, Oxford).
- Hamilton, J.D., 1994a, State-space models, in R Engle, and D McFadden, eds.: *Handbook of econometrics* (North Holland, Amsterdam).
- Hamilton, James D. , 1989b, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica: Journal of the Econometric Society* 57, 357-384.
- Hamilton, James D., and Baldev Raj, 2002, New directions in business cycle research and financial analysis, *Empirical Economics* 27, 149-162.
- Hansen, B. E., 1992, The likelihood ratio test under nonstandard conditions: Testing the markov switching model of gnp, *Journal of Applied Econometrics* 7, S61-S82.
- Hendry, David F., 1995. *Dynamic econometrics* (Oxford University Press, Oxford).
- Judge, George G., 1985. *The theory and practice of econometrics* (Wiley, New York ; Chichester).
- Kim, C.J., and C.R. Nelson, 1999. *State-space models with regime switching* (Massachusetts: MIT Press, Cambridge).
- Kim, Chang-Jin, 1994, Dynamic linear models with markov-switching, *Journal of Econometrics* 60, 1-22.
- Kim, Chang-Jin, and Charles R. Nelson, 2001, A bayesian approach to testing for markov-switching in univariate and dynamic factor models, *International Economic Review* 42, 989-1013.
- Lin, C. F. J., and T. Terasvirta, 1994, Testing the constancy of regression parameters against continuous structural change, *Journal of Econometrics* 62, 211-228.
- Lutkepohl, H., and M Kratzig, 2004. *Applied time series analysis* (Cambridge University Press, Cambridge).
- Luukkonen, Ritva, Pentti Saikkonen, and Timo Terasvirta, 1988, Testing linearity against smooth transition autoregressive models, *Biometrika* 75, 491-499.
- Maddala, G. S., 1977. *Econometrics* (McGraw-Hill, New York ; London).
- Neftçi, Salih N., 1982, Optimal prediction of cyclical downturns, *Journal of Economic Dynamics and Control* 4, 225-241.

- Quandt, R. E., 1984, Computational problems and methods, in Z Griliches, and M. D. Intriligator, eds.: *Handbook of econometrics* (North-Holland, Amsterdam).
- Raj, Baldev, 2002, Asymmetries of business cycle: The markov-switching approach, in A Ullah, A Wan, and A Chaturvedi, eds.: *Handbook of applied econometrics and statistical inference* (Marcel Dekker).
- Terasvirta, T, 1994, Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* 89, 208-218.
- Terasvirta, T, 1998, Modeling economic relationships with smooth transition regressions, in A. Ullah, and D.E. Giles, eds.: *Handbook of applied economic statistics* (Dekker, New York,).
- Tong, Howell, 1983. *Threshold models in non-linear time series analysis* (Springer Verlag, New York).
- Tong, Howell, 1990. *Non-linear time series : A dynamical system approach* (Clarendon Press, Oxford).
- Wooldridge, J. M., 1994, Estimation and inference for dependent processes, in R. F. Engle, and D. L. McFadden, eds.: *Handbook of econometrics* (North-Holland, Amsterdam).

Chapter III A review of the models in the term structure of interest rates

The term structure of interest rates plays a central role in economy, both theoretically and practically. While policy makers conduct monetary policy by shifting interest rates at the short end of yield curve, longer term yields give people's reflection of future economy health. When expectations about future rates change, the yield curve reacts as bonds reprice to reflect these expectation changes. Moreover, longer term yields also compound liquidity risk premium and nonlinear convexity premium. Therefore, the movements in longer end of the term structure reflect not only the variation in expectations but also various risk premiums. Given the importance of interest rates modelling, the related econometric estimation and forecasting methods have received tremendous attention in recent years. In this section, we review various methods used in term structure literature in estimating term premia, from simple regression methods (like Fama and Bliss (1987) and Cochrane and Piazzesi (2005)) to factor models like the Nelson-Siegel factor model and affine term structure models. We stress that the main reason of empirical observation deviating from classical expectations hypothesis is the time-varying term premia. We also utilize the zero-coupon yield data from UK, US and Germany to give examples of different methods in testing and modelling the time-varying term premia.

III.1 Information contained in the term structure of interest rates

Information contained in the term structure of interest rates is valuable to central banks. It is often mentioned that the market expectations implied by the shape of yield curves that provide central banks early warnings of the deterioration of economic conditions. On the reverse side, the short end of the yield curve is under the direct control of central banks, which gives policy makers a powerful tool to achieve their goals in stabilizing economies. From a financial market perspective, the term structure of interest rates is the fundamental building block for many financial instruments. An accurate extraction of information from the yield curve gives market participants advantages in pricing and hedging various financial securities. However, to understand the information contained in the term structure of interest rates, one needs to properly separate the expectations of future movements of the interest rates from the term premia. In other words, a risk-averse investor should distinguish the difference between the actual long rate they required and that implied from a risk-neutral investor.

Recent research has seen the proliferation of studies on explicit modelling of term premia in the bond market. To some extent, this research attention was instigated by Alan Greenspan's speech on yield market "Conundrum" that

"... Long-term interest rates have trended lower in recent months even as the Federal Reserve has raised the level of the target federal funds rate by 150 basis points. This development contrasts with most experience, which suggests that, other things being equal, increasing short-term interest rates are normally accompanied by a rise in longer-term yields... for the moment, the broadly unanticipated behavior of world bond markets remains a conundrum ..."⁶,

in other words, the low long-term interest rates are too low to be properly explained solely by reversions in expectations. Thus, a time-varying term premia has been argued as a potential factor in explaining the puzzling behavior of the long-term interest rates. Then, what is the expectation on interest rates and what are the term premia embedded in the term structure of interest rates?

Classical finance theory states that the long-term bond yield is simply the average of the expected future short-term interest rates. This definition ignores the interest rate risks in the future and therefore is coined as unbiased pure expectation theory of the term structure of interest rates. The term "unbiased" implies that the forward rate is the unbiased estimate of the expected future short rate. However, as we know, unless one holds the long-term bond to its maturity, the holding return on this long-term bond depends on the uncertain future discount rate at time of selling. To compensate the risk of holding long-term bond, investors require term premia on long-term bonds.

III.2 Definitions of term premia

Although progress has been made in the past few years on modelling time-varying term premia, there is relatively little consensus about the multiple definitions of the term premia in empirical works. According to Shiller and McCulloch (1990), there are three popular definitions of the term premia in term structure of interest rates literature, which are stated below.

⁶ Testimony of federal Reserve Board Chairman Alan Greenspan to the U.S. Senate, February 16th, 2005

First is the return term premia, which is defined as the difference between conditional expected return on holding a long-term bond for one period and the one-period short rate.

$$\Lambda_t^{(n),r} = E_t \left(r_{t+1}^{(n)} \right) - y_t^{(1)} \quad (\text{III.1})$$

On the right-hand-side of (III.1), the return on holding a long-term bond for one period is defined as $r_{t+1}^{(n)} = \log \left(P_{t+1}^{(n-1)} \right) - \log \left(P_t^{(n)} \right)$, where $P_{t+1}^{(n-1)}$ is the price of the n -period zero-coupon bond at time $t+1$ (time to maturity for this n -period bond has reduced to $n-1$). $y_t^{(1)} = -\log P_t^{(1)}$ is the one-period yield, i.e. the short rate for one period. On the left-hand-side, $\Lambda_t^{(n),r}$ is the one-period holding-period return term premia (the multi-period holding-period return is similar to the one-period case).

The second definition of term premia, among the three, is called the forward term premia, which is defined as the difference between forward rate and the expectation of the corresponding future spot rate. The mathematical expression is as shown in (III.2),

$$\Lambda_t^{(n),f} = f_t^{(n)} - E_t \left(y_{t+n-1}^{(1)} \right) \quad (\text{III.2})$$

where $f_t^{(n)} = \log \left(P_t^{(n-1)} \right) - \log \left(P_t^{(n)} \right)$ is the one period forward rate at time t , $y_{t+n-1}^{(1)}$ is the future ($n-1$ period ahead) one-period spot rate, and $\Lambda_t^{(n),f}$ is the forward term premia.

The last definition of the term premia refers to yield term premia, which is the difference between the yield on a zero-coupon bond and the average of the conditional expected future short rates to the maturity of the zero-coupon bond:

$$\Lambda_t^{(n),y} = y_t^{(n)} - \frac{1}{n} \sum_{i=0}^{n-1} E_t \left(y_{t+i}^{(1)} \right) \quad (\text{III.3})$$

where $y_t^{(n)} = -\frac{1}{n} \log \left(P_t^{(n)} \right)$ is the yield on a n period zero-coupon bond, and $y_{t+i}^{(1)}$ is the one-period yield at time $t+i$.

Given the three popular definitions of the term premia in the term structure of interest rate literatures, one can easily show that those three definitions are essentially the same. For example, from (III.2), it follows that $f_t^{(1)} = E_t \left(y_t^{(1)} \right) + \Lambda_t^{(1),f}$, $f_t^{(2)} = E_t \left(y_{t+1}^{(1)} \right) + \Lambda_t^{(2),f}$, \dots ,

$f_t^{(n)} = E_t \left(y_{t+n-1}^{(1)} \right) + \Lambda_t^{(n),f}$. After substituting these expected i -period later yield into (III.3), we have

$$y_t^{(n)} = \frac{1}{n} \left[f_t^{(1)} + f_t^{(2)} + \dots + f_t^{(n)} - \left(\Lambda_t^{(1),f} + \Lambda_t^{(2),f} + \dots + \Lambda_t^{(n),f} \right) \right] + \Lambda_t^{(n),y} \quad (\text{III.4})$$

from which the yield term premia can be expressed as the average of the forward term premia:

$$\begin{aligned} \Lambda_t^{(n),y} &= \frac{1}{n} \sum_{i=1}^n \Lambda_t^{(i),f} + y_t^{(n)} - \frac{1}{n} \sum_{i=1}^n f_t^{(i)} \\ &= \frac{1}{n} \sum_{i=1}^n \Lambda_t^{(i),f} + y_t^{(n)} - \frac{1}{n} \sum_{i=1}^n \left[\log \left(P_t^{(i-1)} \right) - \log \left(P_t^{(i)} \right) \right] \\ &= \frac{1}{n} \sum_{i=1}^n \Lambda_t^{(i),f} + y_t^{(n)} - \frac{1}{n} \left[\log \left(P_t^{(0)} \right) - \log \left(P_t^{(n)} \right) \right] \\ &= \frac{1}{n} \sum_{i=1}^n \Lambda_t^{(i),f} - \frac{1}{n} \log \left(P_t^{(n)} \right) + \frac{1}{n} \log \left(P_t^{(n)} \right) \\ &= \frac{1}{n} \sum_{i=1}^n \Lambda_t^{(i),f} \end{aligned} \quad (\text{III.5})$$

The relationship between the forward term premia and the holding-period return term premia is given by Cochrane and Piazzesi (2008) as

$$\Lambda_t^{(n),f} = \Lambda_t^{(n),r} + E_t \left(\Lambda_{t+1}^{(n-1),r} - \Lambda_t^{(n-1),r} \right) + \dots + E_t \left(\Lambda_{t+n-2}^{(2),r} - \Lambda_{t+n-3}^{(2),r} \right) \quad (\text{III.6})$$

One can see from above that in fact the three definitions of term premia are equivalent, in the sense that if one specification holds with constant term premia, the remaining two specifications of term premia are also holding for constant term premia.

III.3 Regression analysis of term premia

Independent of which the above term premia is chosen to be estimated, the separation between the unobserved expectation component and term premia is always a challenging work. However, one can always start with simple regression analysis to test the consistency of the celebrated expectations hypothesis.

III.3.1 Testing on forward term premia

Under the assumption of rational expectations, the difference between current forward rate and the ex post realised short rate should not be predictable using ex ante variables. Violation of the above implies either an inefficiency of the rational expectations hypothesis or the existence of a forward term premia. By adopting the validation of the rational

expectations hypothesis, the forward term premia can be measured as the predictable component of the difference between forward rate and ex post short rate. Subtracting the realised one-period yield at time $t+n-1$ from both sides of (III.2), yields

$$f_t^{(n)} - y_{t+n-1}^{(1)} = \left[E_t \left(y_{t+n-1}^{(1)} \right) - y_{t+n-1}^{(1)} \right] + \Lambda_t^{(n),f} \quad (\text{III.7})$$

where the first term on the right-hand side of (III.7) is the expectations error, which should be zero on average. One possible test on the expectations hypothesis would be by constructing the following regression equation

$$f_t^{(n)} - y_{t+n-1}^{(1)} = \alpha + \sum_{i=1}^P X_{i,t}' \beta_i + \varepsilon_t \quad (\text{III.8})$$

where α is the constant term, $X_{i,t}$ is the forecasting variables with coefficients β_i s. Given that the expectation error is zero under rational expectations (that is, over a long time period, the downward biased expectation of short rate would be balanced by the upward biased ones),

$\alpha + \sum_{i=1}^P X_{i,t}' \beta_i$ constitutes the forward term premia. If we rearrange (III.8) by replacing the second term on the right-hand side with $\beta \left(f_t^{(n)} - y_t^{(1)} \right)$, the above equation is equivalent to Fama and Bliss (1987) regression equation which takes the form of

$$\begin{aligned} f_t^{(n)} - y_{t+n-1}^{(1)} &= y_{t+n-1}^{(1)} - y_t^{(1)} + \alpha + \beta \left(f_t^{(n)} - y_t^{(1)} \right) \\ &\Rightarrow \\ y_{t+n-1}^{(1)} - y_t^{(1)} &= -\alpha + (1 - \beta) \left(f_t^{(n)} - y_t^{(1)} \right) + \varepsilon_t \\ &= \alpha^* + \beta^* \left(f_t^{(n)} - y_t^{(1)} \right) + \varepsilon_t \end{aligned} \quad (\text{III.9})$$

where $\alpha^* = -\alpha$ and $\beta^* = (1 - \beta)$. Note that, under pure expectations hypothesis, $\alpha^* = 0$ and $\beta^* = 1$, which implies that the perfect forecast of future one-period short rate - $y_{t+n-1}^{(1)}$, is the one-period forward rate $f_t^{(n)}$. A rejection of $\beta^* = 1$ is interpreted as the evidence of a time-varying forward term premia. Fama and Bliss, using (III.9), found that the spread between the n-year forward rate and the one-year yield $\left(f_t^{(n)} - y_t^{(1)} \right)$ predicts the one year excess return of the n-year bond $\left(y_{t+n-1}^{(1)} - y_t^{(1)} \right)$, with R^2 about 18%.

III.3.2 Testing on long-term premia

Long-term premia are usually associated with the holding-period return premia on long-term bonds. Similar to the forward term premia regressions aforementioned, the excess holding-period returns on long-term bonds should not be predictable under the joint assumption of the expectations hypothesis and rational expectations. If, in fact the excess return is predictable, there exists a holding-period term premia.

Cochrane and Piazzesi (2005) forecast one-year holding returns of the bonds with 2-5 years of maturities using all forward rates, not just the matched forward spot spreads as in Fama-Bliss regressions. The regressions studied in their paper takes the following form

$$r_{t+1}^{(n)} - y_t^{(1)} = b_0^{(n)} + b_1^{(n)} y_t^{(1)} + b_2^{(n)} f_t^{(2)} + \dots + b_5^{(n)} f_t^{(5)} + \varepsilon_{t+1}^{(n)} \quad (\text{III.10})$$

where the left-hand-side variable is the excess returns of bonds with different maturities over the one-period short rate; and on the right-hand side is a linear combination of the forward rates. Results based on (III.10) reported the high predictability of the linearly combined single factor which gives R^2 as high as 40%. In addition, this single return forecasting factor is unrelated to the movements of conventional factors (level, slope and curvature factors) that are suggested by finance theory, but it is countercyclical and forecasts stock returns. Their paper examined only one to five year maturity bonds at a one year horizon in the US market; however, it is interesting to see if this tent-shaped single factor appears in other markets for different maturity bonds with longer return horizons. In section III.3.4, we utilize Cochrane and Piazzesi (2005)'s model on German and UK's zero coupon bond data to see if we still could get similar results as in their study for the US market. We also extend their original model to consider more maturities of the bond yields.

III.3.3 The Campbell-Shiller (1991) puzzle

Campbell and Shiller (1991) obtained self-contradictory results, while regressing short-term changes of long-term rates and long-term changes of short-term rates, respectively, on the slope of the yield curve. If we expect rational expectations to hold, the distribution of the difference between conditional value of ex ante variable $E_t(y_{t+i}^{(1)})$, in (III.3), and the ex post realised value of $y_{t+i}^{(1)}$ will have a zero mean, which is equivalent to say that under the hypothesis of perfect foresight, $E_t(y_{t+i}^{(1)}) = y_{t+i}^{(1)} + v_t^i$, where v_t^i is an i.i.d. process with zero mean

and variance of σ_i^2 . Hence, we could rewrite (III.3) by subtracting $y_t^{(1)}$ from both sides and parameterizing it as

$$\frac{1}{n} \sum_{i=0}^{n-1} y_{t+i}^{(1)} - y_t^{(1)} = \alpha + \beta \left(y_t^{(n)} - y_t^{(1)} \right) + e_t^{(n)} \quad (\text{III.11})$$

where $e_t^{(n)} = \frac{1}{n} \sum_{i=0}^{n-1} v_{t+i}$, and we can test on the null hypothesis of $\beta=1$ to claim that any violation to this is regarded as a violation of the expectations hypothesis.

By projecting the short-term changes of long-term rates on the slope of the yield curve, we can rewrite (III.1) as

$$\begin{aligned} E_t \left[\log(P_{t+1}^{n-1}) - \log(P_t^n) \right] &= y_t^{(1)} + \Lambda_t^{(n),r} \\ \Rightarrow \begin{cases} E_t \left[-(n-1) y_{t+1}^{(n-1)} + n y_t^{(n)} \right] &= y_t^{(1)} + \Lambda_t^{(n),r} \\ -(n-1) E_t \left(y_{t+1}^{(n-1)} \right) + n y_t^{(n)} &= y_t^{(1)} + \Lambda_t^{(n),r} \\ (n-1) E_t \left(y_{t+1}^{(n-1)} \right) &= n y_t^{(n)} - y_t^{(1)} - \Lambda_t^{(n),r} \end{cases} \end{aligned} \quad (\text{III.12})$$

Subtracting $(n-1) y_t^{(n)}$ from both sides of the last equation above, we obtain that

$$(n-1) E_t \left(y_{t+1}^{(n-1)} - y_t^{(n)} \right) = \left(y_t^{(n)} - y_t^{(1)} \right) + \Lambda_t^{(n),r} \quad (\text{III.13})$$

Again, assuming the rational expectation hold, (III.13) can be re-parameterized as

$$y_{t+1}^{(n-1)} - y_t^{(n)} = \frac{\gamma}{n-1} \left(y_t^{(n)} - y_t^{(1)} \right) + e_t^{(n)} \quad (\text{III.14})$$

where $\gamma \neq 1$ will be tested against the null hypothesis that the expectation hypothesis does hold ($\gamma=1$).

By estimating (III.11), Campbell and Shiller noted that β coefficients are positive and, in most of the cases, are significantly different from zero that clearly reject the expectations hypothesis. In addition, the adjusted R^2 suggests that the slope of the yield has some explaining power for the long-term changes of short rates. The estimates of γ coefficients in (III.14), however, are negative (wrong sign, despite the rejection of

expectation hypothesis) and the adjusted R^2 suggests that the slope of the yield curve explains nothing on the short-term changes of long-term rates. As noted in Campbell (1995):

“ ... (The results) seems self-contradictory ... that yield spreads don't forecast short-run changes in long yields contradict the expectation hypothesis of the term structure. Yet... yield spreads do forecast long-run changes in short yields offer some support for expectation hypothesis, at least at very short and very long maturities. What are we to make of these mixed results? ...”

Attempts to resolve the Campbell and Shiller puzzle have been made mainly on four directions: a time-varying term premia, irrational expectations, measurement error problems in statistical tests, and decline in predictability of short-term rate.

A time-varying term premia, if positively correlated with the slope of yield curve, can be shown that is able to generate downward biased estimates of β and γ . However, the inability of generating negative estimates of γ has been argued in several studies (e.g. Hardouvelis (1994), Thornton (2006)), and suggest that a more plausible resolution would be the irrational expectations hypothesis of Campbell and Shiller (1991) and Froot (1989).

The irrational expectation explanation asserts that the downside biased estimates of β and γ can be explained by the overreaction of long-term rates to the short-term rates. In the case when expectations of the future short-term rate rise, the long-term rate increases by an amount that is larger than the conventional expectations hypothesis would support, causing the slope of the yield curve to steepen. As a result, when markets correct their irrational expectations bias over time, the long-term rate falls as the short-term rate rises. Therefore, this creates the negative correlation between yield curve slope and both the short-term changes in long-term rate and the long-term changes in short-term rate.

Campbell (1995) imputes the failure of the expectations hypothesis to the changing rational expectations on long-term bond returns. Since the slope of the yield curve (yield spread) is the return on a long-term bond minus the return on a short-term bond, a positive slope, then, either represents that the long-term bond is expected to have a higher return than a series of short-term bonds over its life, or that the short-term rate is expected to rise over the life of the long-term bond, or the mixed result of the above two. However, the expectations

hypothesis rules out the possibility that investors would rationally expect unusual high or low returns on long-term bond, the rising short-term rate is then the only outcome of a positively sloped yields spreads. As a result, the β and γ coefficients are biased downwards when compared to the general case that yields spreads reflecting both the changing rational expectations of excess return of long-term bond over short-term bond and the changing rational expectations of future short-term rate. Campbell further suggests that the changing rational expectations on the excess long-term bond return can be viewed as a measurement error that affects positively the right-hand-side of (III.14) but negatively on the left-hand-side, thus, causing changing signs of γ from positive to negative. In contrast, this measurement error only affects negatively the right-hand-side of (III.11), which could change the size of β but not the sign of it.

Closely related with Campbell's explanation on the failure of expectations hypothesis, the decline in predictability of the short-term rate, which was first pointed out by Mankiw and Miron (1986). It is argued that the commitment of central banks to smooth interest rate in medium term explains the random walk behaviour of the short-term rate after the foundation of the Fed in 1913. Furthermore, Balduzzi, Bertola and Foresi (1997) suggests that central bank's effort to make both very short-term and very long-term rates predictable explains why the β coefficient is less downward biased in both short-term and long-term.

III.3.4 Fama-Bliss (1987) and Cochrane-Piazzesi (2005) regression analysis

In this empirical study subsection, we test the forward term premia of Fama and Bliss (1987) and long term return term premia of Cochrane and Piazzesi (2005). Specifically, we run regressions of the forms as in (III.9) and (III.10), using zero-coupon bond data from US, UK and Germany for comparisons⁷.

Table III-2 (in Appendix III) shows the descriptive statistics of the yield curves and calculates the average excess holding period returns for each country. The expectations hypothesis seems to be held. The slight upward sloping average yield curve, which is often argued as the presence of a small "liquidity premia", translate to a higher average excess holding period return for longer maturity bonds, albeit with increasing volatility at the long-end. From the table, one can also spot the low Sharp Ratio of bonds when compared with

⁷ The zero-coupon bond data (Fama-Bliss data) for the US are obtained from John Cochrane's website: <http://faculty.chicagobooth.edu/john.cochrane/research/papers/>. The data for the UK and Germany are obtained from Bank of England's and Bundesbank's websites.

stocks. This characteristic, which is common in treasury bonds, reflects the small market price of risk that assigned to the volatility factors influencing treasury bond returns.

Following Fama and Bliss (1987), we forecast the changes of short rate by using the *Forward-Short rate-Spread (FSS)* as in (III.9). The "complementary regression", which is defined as $r_{t+1}^{(n)} - y_t^{(1)} = a^C + b^C (f_t^{(n)} - y_t^{(1)})$ in their paper, is also implemented for comparison. The results are shown in Table III-3. Panel A shows the *FSS* does a pretty good job in forecasting the future short rate changes in the UK and German markets but not in the US market. The expectations hypothesis, if holding, predicts a coefficient of 1 on b^* , since the forward rate should be a perfect forecast of the future spot rate. Specifically, if the one-period forward rate $f_t^{(n)}$ is 1% higher than the current short rate, the expectations hypothesis suggests a one-for-one increase in the spot rate one period from now. We see the UK and German markets generally follow this prediction. The US market, however, shows no predictive power of the *FSS*. Particularly, the coefficients of b^* is close to zero and the adjusted R^2 is low at 1-year horizon. Although the coefficient of b^* increases to 0.879 at the 5-year horizon, the story seems to be that the *FSS* forecasts poorly on the future short rate changes but well on the excess holding period returns in the short run. As stated in Cochrane and Piazzesi (2005), the fact that the yield changes in short run are almost unpredictable is mechanically equivalent to saying bond returns are predictable in the short run. If we rewrite the "complementary regression" as

$$\begin{aligned}
 r_{t+1}^{(n)} - y_t^{(1)} &= a^C + b^C (f_t^{(n)} - y_t^{(1)}) \\
 &\Leftrightarrow \\
 p_t^{(n-1)} - p_{t+1}^{(n)} + p_t^{(1)} &= a^C + b^C (p_t^{(n-1)} - p_t^{(n)} + p_t^{(1)}) \\
 -p_{t+1}^{(n)} &= a^C - (1 - b^C)(p_t^{(n-1)} + p_t^{(1)}) - b^C p_t^{(n)} \\
 ny_{t+1}^{(n)} &= a^C - (1 - b^C)(p_t^{(n-1)} + p_t^{(1)}) + b^C y_t^{(n)},
 \end{aligned} \tag{III.15}$$

the implication of $a^C = 0$ and $b^C = 1$ is that the bond prices and bond yields follow random walk (completely unpredictable). However, a^C and b^C for the US data in Panel B are not 0 and 1 exactly, and yields are stationary and moving slowly over time. The sluggish changes in the yields make the changes in the yields in short run unpredictable but, as the time horizon increases, the predictability rises. The building up of the b^* coefficients and the \bar{R}^2 as n

increases, on the other hand, suggests the expectations hypothesis in the long run is held in the US market.

In comparison with the US market, the UK and German markets show support for the expectations hypothesis, even in the short run. The close to 1 b^* coefficient indicates a quicker adjustment of the short rate in a short time horizon. Even more, the \bar{R}^2 builds up to more than 50% in the long run. In contrast to this huge predictability in the changes of yields, the bond returns are completely unpredictable. Not only the b^c coefficients are eventually zero, the \bar{R}^2 s are even negative for some time horizons.

If the *FSS* indeed forecast excess bond returns but not changes in yields in the US market while the reverse is correct in the UK and German market, the next question raised naturally is what about other factors as the right-hand-side variables? Should other forward rates also forecast excess bond return in the US market and can we predict the excess bond returns using more forward rates in the UK and German markets? Cochrane and Piazzesi (2005) uses all forward rates to forecast excess bond returns with 1-year holding period. Following their procedures, we extend their study to the UK and German markets. Additionally, we also forecast the excess bond returns which have a 2-year holding period.

Panel B in Table III-4 shows the results on restricted and unrestricted regressions of the 1-year excess bond returns on all forward rates for each country. The unrestricted regression takes the form of (III.10), where the excess bond returns are forecasted by a linear combination of all the forward rates. The restricted regressions project the excess bond returns on a single factor as shown in

$$r_{t+1}^{(n)} - y_t^{(1)} = \beta^{(n)} \left(\gamma_0 + \gamma_1 y_t^{(1)} + \sum_{k=2}^K \gamma_k f_t^{(k)} \right) + \varepsilon_{t+1}^{(n)}, \quad (\text{III.16})$$

where K is the number of forward rates included and $\beta^{(n)}$ is restricted as $\frac{1}{K-1} \sum_{n=2}^K \beta^{(n)} = 1$.

The single factor (SF_t), which is the common combination of forward rates

$SF_t = \left(\gamma_0 + \gamma_1 y_t^{(1)} + \sum_{k=2}^K \gamma_k f_t^{(k)} \right)$, tells us the common movements of all the excess bond returns

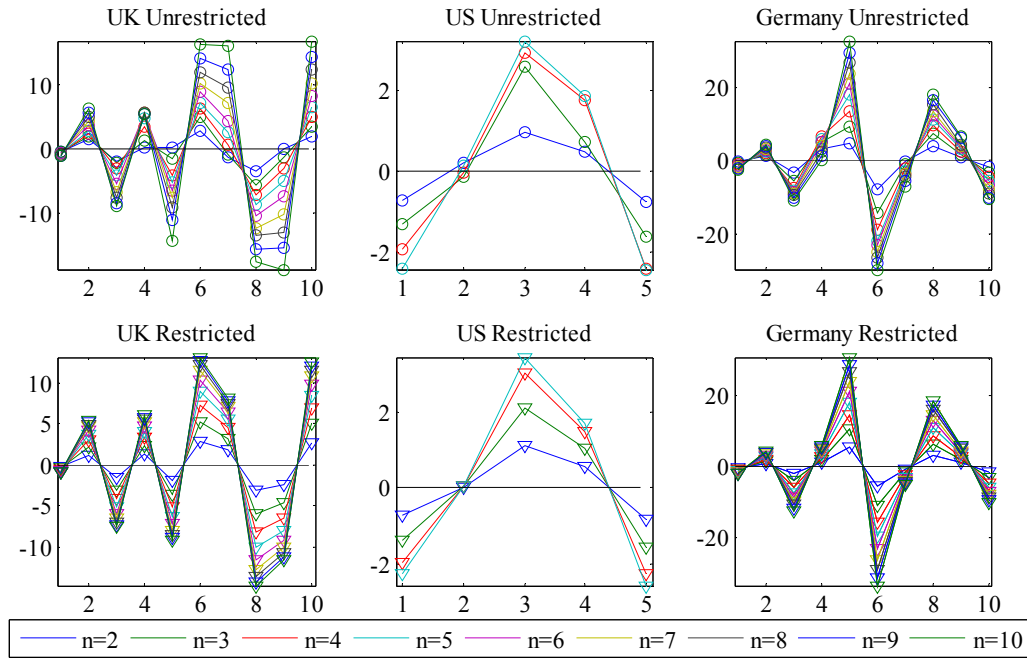
at a given date. It is the fitted value of a regression from projecting the average excess bond returns on all forward rates. The average excess bond return regression takes the form as

$$\frac{1}{K-1} \sum_{n=2}^K \left(r_{t+1}^{(n)} - y_t^{(1)} \right) = \gamma_0 + \gamma_1 y_t^{(1)} + \sum_{k=2}^K \gamma_k f_t^{(k)} + \varepsilon_{t+1}. \quad (\text{III.17})$$

The predictability of the combination of all forward rates, which is measured by \bar{R}^2 as shown in the unrestricted regressions, suggests a significant improvement from just using the *FSS* as the sole predictor. The \bar{R}^2 in the US market is up from 6.2%-14% to 24%-28%, while the improvements in the UK and German market are also significantly rising from -0.2%-0.7% to 4%-8.7% and 2.3%-5.4% to 10.7%-16.1%, respectively. The \bar{R}^2 in the restricted regressions are almost unaffected, and the plots of those restricted coefficients in Figure III-1 shows great resemblances to the unrestricted coefficients. Although the shape of those coefficients in the UK and German markets do not show the usual “tent” shape found in the US market, one can easily spot the common movements of all the forward rates. Panel A of Table III-4 shows the regression results of (III.17), which is used to form the single factor that is used in the restricted regressions. The \bar{R}^2 statistics show the single factor can forecast up to 26.1% (14.9%) of the variation in the average excess bond returns in the US (German) market. The statistic in the UK market, however, is much lower with only 4.6% of the variation can be forecasted using forward rates. The χ^2_{10} statistic also indicates the joint insignificance of the right-hand-side variables in (III.17) for the UK market.

Our finding that the shapes of those coefficients in the UK and German markets do not show the similar “tent-shaped” coefficients found in the US market is also confirmed by Kessler and Scherer (2009). This provides us some evidence that the results of Cochrane and Piazzesi (2005) are not transferable to other international bond markets and time frames than what they have studied. One possible explanation for the missing pattern of the estimated coefficients across the UK and German market, as mentioned in Kessler and Scherer (2009), is due to the high correlations among the explaining variables (i.e. the forward rates with various maturities) in the UK and German market. Even if there would be economic factors driving the impact of the various forward rates on bond returns, the apparent high correlations among those forward rates would prevent us from identifying them robustly. In addition, the lack of consistent pattern in the estimates of the forward rate parameters across the UK and German markets reveals that there is no systematic and uniform effect of the single factor on forecasting the bond returns. The increases in \bar{R}^2 in all three markets, however, indicate that there is a strong relationship between the forward rates curve and the bond returns.

Figure III-1: Unrestricted and restricted coefficients



The time series of the single factors (C-P forecast) in different markets are plotted in Figure III-2 to Figure III-4 below. The low variation of the single factor in the UK market is consistent with the finding of insignificant γ coefficients in Panel A of Table III-4. Comparatively, the F-B forecast (average of Fama-Bliss forecast) is even worse, shows no co-movement with the average realized bond return. The single factors in the US and German markets, on the opposite, show much stronger co-movements with the realized bond returns.

In summary, the ability of the linear combination of all forward rates to predict excess bond returns in the US and German market advocate the failure of expectations hypothesis. Under the expectations hypothesis, the n -year forward rate should be the optimal forecast of the spot rate $n-1$ years hence. The Fama-Bliss regression shows the expectation hypothesis only holds in the long run but not the short run. Specifically, the *FSS* cannot forecast the future spot rate changes in the short run but able to tell the variation of the excess bond returns. The C-P regression further improves on the Fama-Bliss's method by allowing more forward rates to forecast excess bond returns, and shows impressive predictability in the US and German market. Although this predictability is weak in the UK market, the comparative advantage over the Fama-Bliss regression is substantial.

Figure III-2: Forecast and realized average excess bond return in the UK⁸

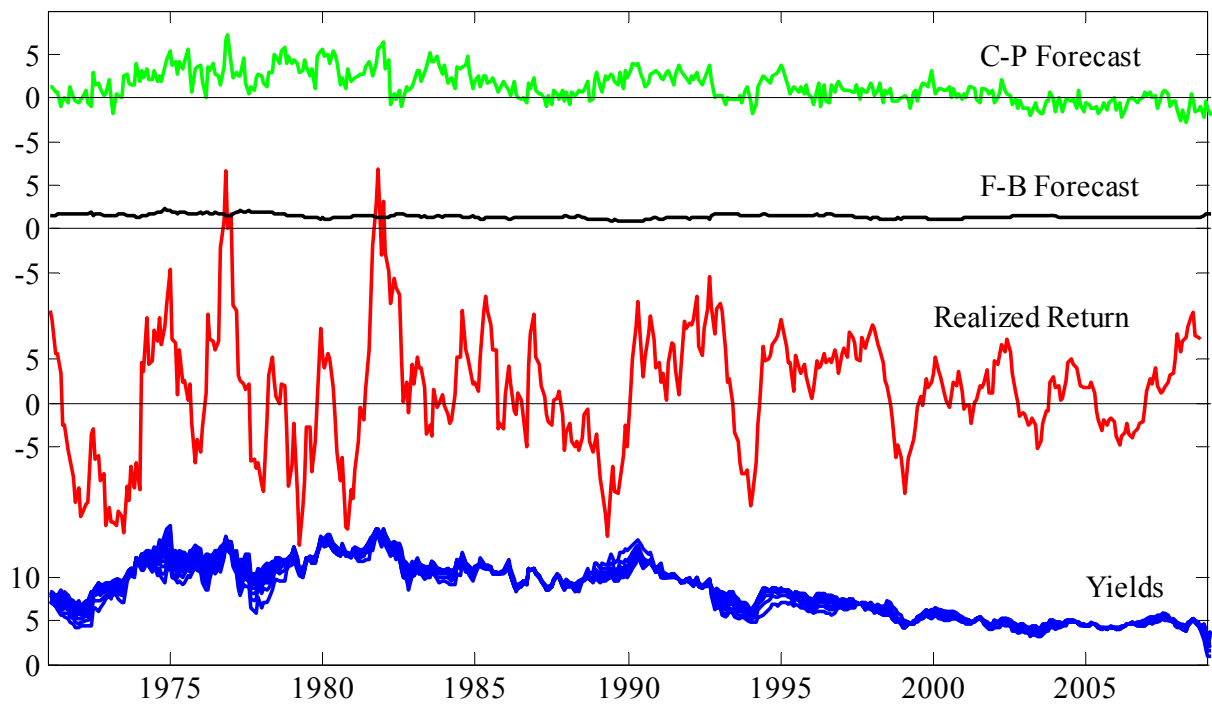
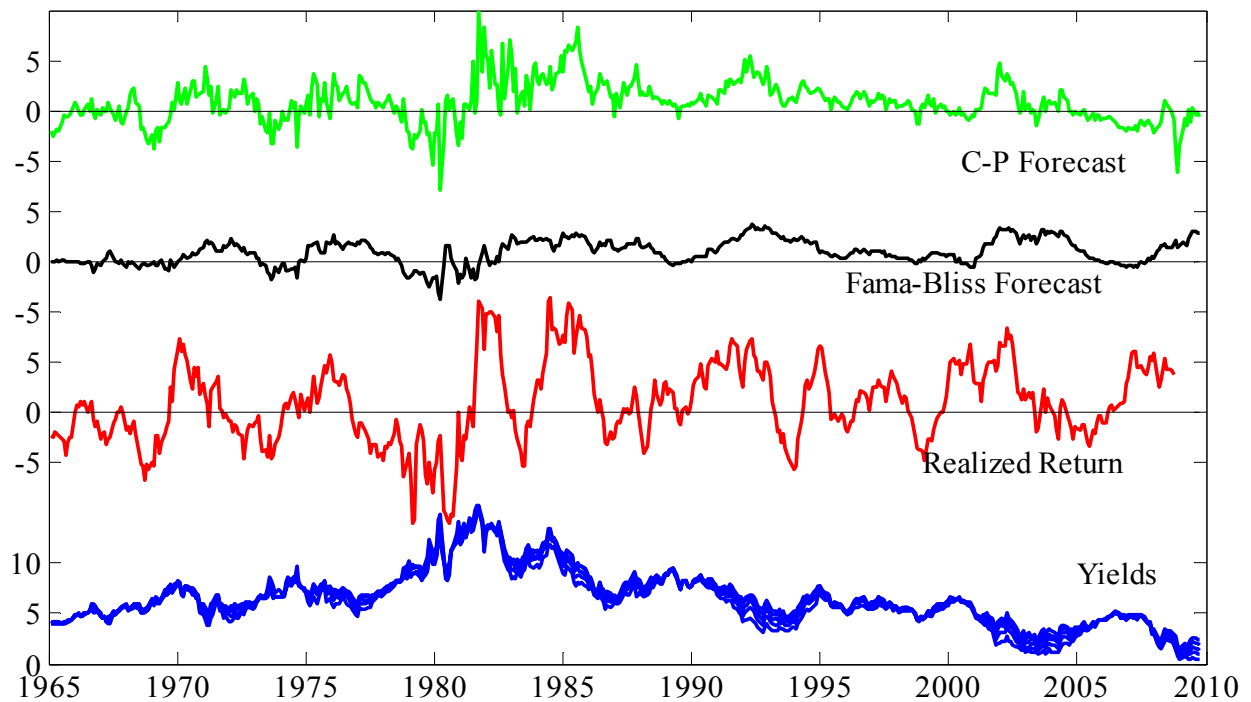
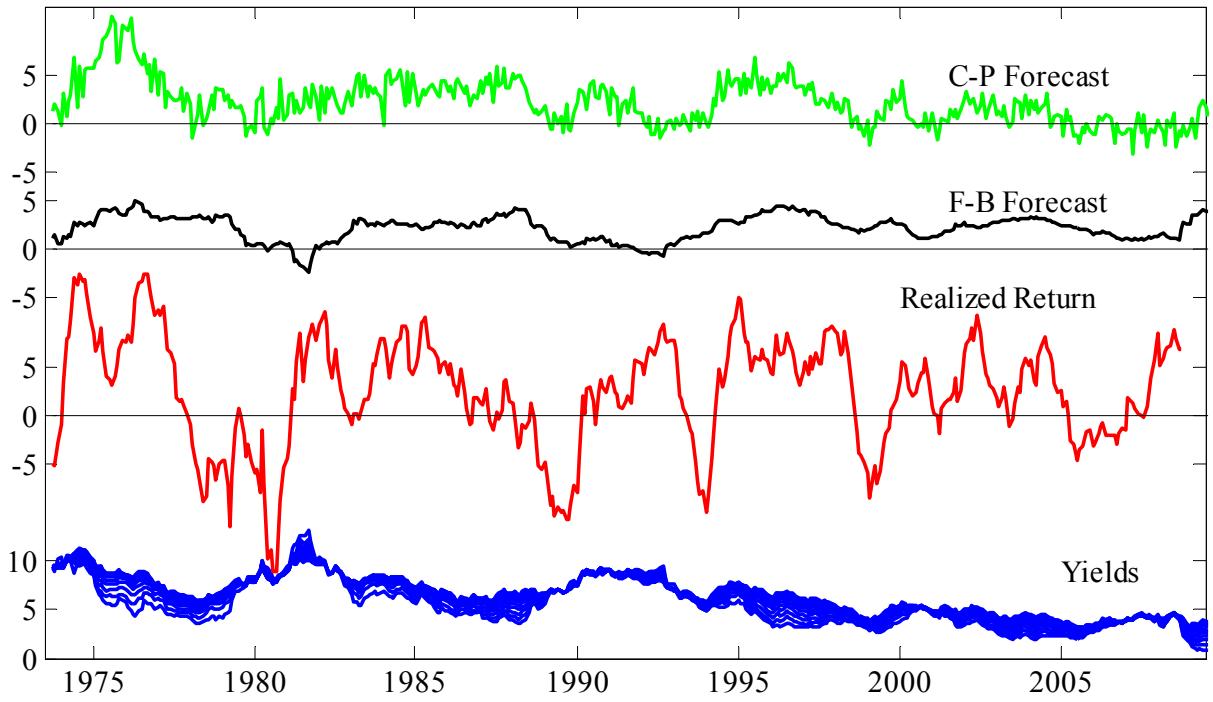


Figure III-3: Forecast and realized average excess bond return in the US



⁸ The realized average excess bond return lines plotted in Figure III-2 to Figure III-4 are shifted back one year.

Figure III-4: Forecast and realized average excess bond return in German



Given the impressive forecasting power of all the forward rates on the excess bond returns with 1-year holding period, the next question naturally arising is "Can we find the similar pattern for a 2-year holding period?". Since the annual excess bond return with a 2-year holding period is $r_{t \rightarrow t+2}^{(n)} - y_t^{(1)} = \frac{(r_{t+1} + r_{t+2})}{2} - y_t^{(1)}$, we can infer the properties of $r_{t \rightarrow t+2}^{(n)}$ by studying on r_{t+2} . The 2-year unrestricted regression of r_{t+2} takes the form of

$$r_{t+2}^{(n)} - y_t^{(1)} = \tilde{b}_0^{(n)} + \tilde{b}_1^{(n)} y_t^{(1)} + \sum_{k=2}^K \tilde{b}_k^{(n)} f_t^{(k)} + \varepsilon_{t+2}^{(n)}. \quad (\text{III.18})$$

To form the 2-year constricted regressions, firstly, we need to forecast the single factor at time $t+1$, which is

$$SF_{t+1} = \tilde{\gamma}_0^{SF} + \tilde{\gamma}_1^{SF(1)} y_t^{(1)} + \sum_{k=2}^K \tilde{\gamma}_k^{SF} f_t^{(k)} + \varepsilon_{t+1}^{SF}. \quad (\text{III.19})$$

The fitted value of SF_{t+1} , which we denote it as $\widehat{SF}_{t+1} = \tilde{\gamma}_0^{SF} + \tilde{\gamma}_1^{SF(1)} y_t^{(1)} + \sum_{k=2}^K \tilde{\gamma}_k^{SF} f_t^{(k)}$, is used to form the restricted regression as

$$r_{t+2}^{(n)} - y_t^{(1)} = \tilde{\beta}^{SF,(n)} \widehat{SF}_{t+1} + \varepsilon_{t+2}^{SF,(n)}. \quad (\text{III.20})$$

Table III-5 presents the results of the 2-year excess bond return regressions. The \bar{R}^2 statistics in Panel A suggest that the linear combination of all forward rates at time t can explain 45.6% (37.4% and 42.2%) of the single factor's variation two years hence in the UK (the US and German) market. This high explanation power of the forwards rates on the single factor is not a surprise, since the single factor at time t (SF_{t+1}) is a linear combination of all the forward rates at time t in (III.17). This creates a mapping of coefficients from $\tilde{\gamma}_0^{SF}, \tilde{\gamma}_1^{SF}, \dots, \tilde{\gamma}_k^{SF}$ to $\gamma_0, \gamma_1, \dots, \gamma_k$ as

$$\begin{aligned} SF_{t+1} &= \tilde{\gamma}_0^{SF} + \tilde{\gamma}_1^{SF} y_t^{(1)} + \sum_{k=2}^K \tilde{\gamma}_k^{SF} f_t^{(k)} + \varepsilon_{t+1}^{SF} \\ &= \gamma_0 + \gamma_1 y_t^{(1)} + \sum_{k=2}^K \gamma_k f_t^{(k)} + \varepsilon_{t+1}^{SF} \\ &= SF_t + \varepsilon_{t+1}^{SF}. \end{aligned}$$

If the single factor follows an AR(1) process, we should observe the $\tilde{\gamma}_0^{SF}, \tilde{\gamma}_1^{SF}, \dots, \tilde{\gamma}_k^{SF}$ have a similar pattern as in $\gamma_0, \gamma_1, \dots, \gamma_k$. In other words, we can forecast the future movements of the single factor from its history. We plot these two sets of coefficients in the middle panels in Figure III-5 to Figure III-7. In the UK market, the two sets of coefficients on the 1 to 5 year forward rates are not walking in step with each other but improve in longer maturities. For the US market, we see a constant level shifts from one to another on the middle maturities, but opposite signs on the very short and long maturities. In the German market, we find strong resemblance between the two sets of coefficients, which suggests the single factor would not only forecast the excess bond returns with 1-year holding period but also those with a 2-year holding period. This is consistent with the findings in Panel B of Table III-5, where the results of the unrestricted and restricted regressions for each market are presented. We see the \bar{R}^2 statistics of both unrestricted and restricted 2-year regressions in the UK and the UK market declined substantially when compared to the 1-year's counterparts. However, the 2-year unrestricted regressions in the German market still obtain relatively high \bar{R}^2 statistics that varies from 12.5% to 17.4%.

Figure III-5: UK 2-year holding period regression coefficients

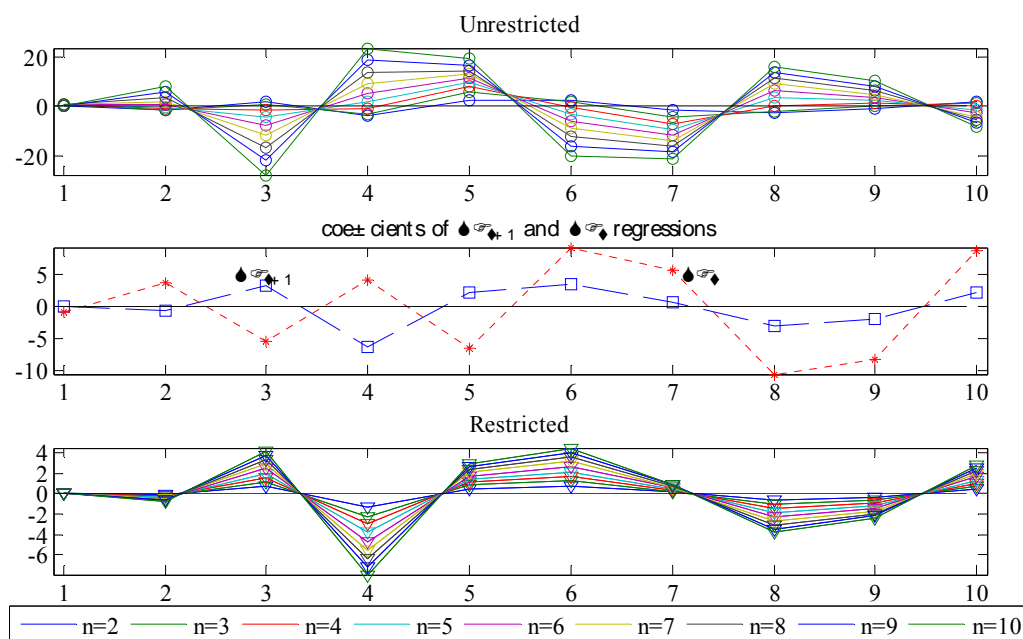


Figure III-6: US 2-year holding period regression coefficients

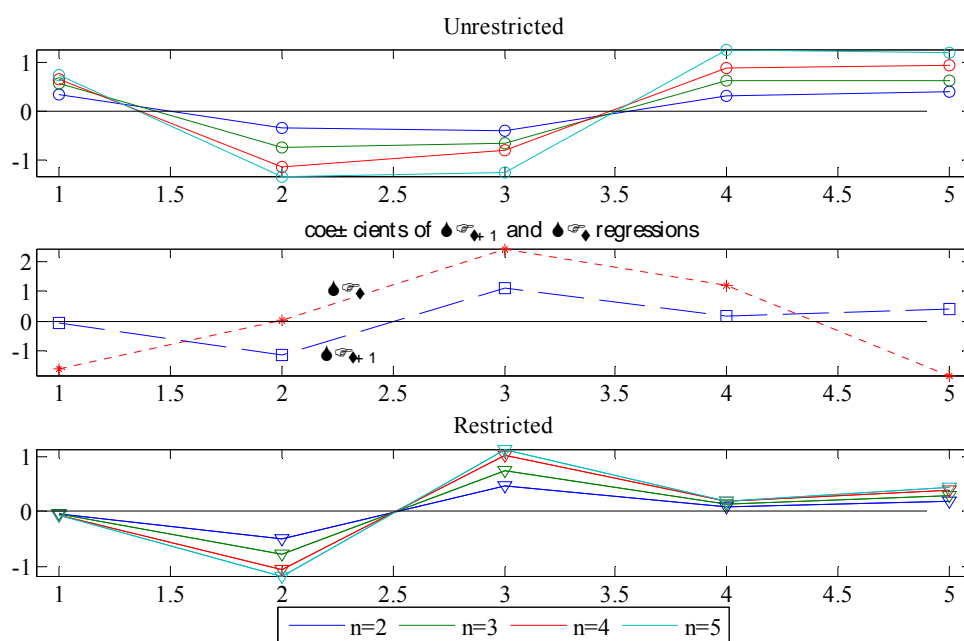
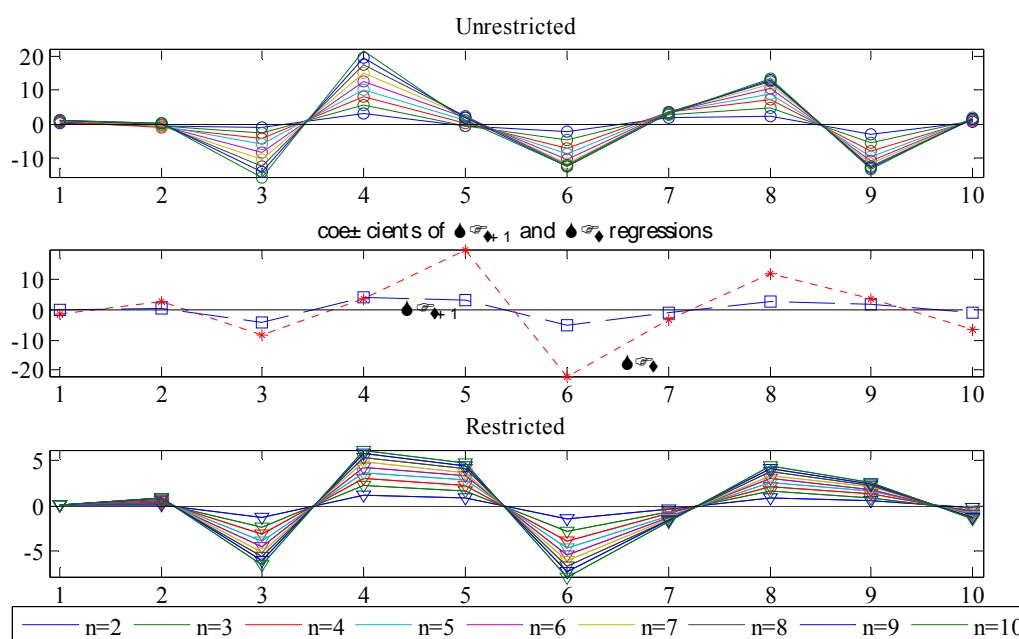


Figure III-7: German 2-year holding period regression coefficients



To summarize, the forecasting exercises in this section show, in general, the failure of the expectations hypothesis. Given this, we follow Cochrane and Piazzesi (2005) to forecast the excess bond returns using all forward rates in the UK, the US and German. The single factor that formed by a linear combination of all forward rates, explains a substantial amount of the variation in excess bond returns with 1-year holding period in the US and German markets but not in the UK market. Only the German market has seen a non-diminishing power of the single factor in forecasting the excess bond returns with a 2-year holding period. The single factor, as advocated in Cochrane and Piazzesi (2005), is correlated with business cycle and suggests a business cycle related risk premia. This reminds us that, apart from the aforementioned regression methods on testing expectations hypothesis, the failure of the expectations hypothesis has also been studied extensively using affine term structure models in recent years. What these affine model based researches argue is that the variation in excess returns of bonds can only be exploited by taking on risks. It is possible that a dynamic investment strategy produces positive returns on average by taking long position in long-term bond when yield curve is steep and negative position when yield curve is flat, but only if the positive excess return is simply the compensation for the risks underlying the strategy. In later sections, we show how affine term structure models can be used to identify the source of

variation in risks, and what the implications of those models on the dynamics of term structure of interest rates are. But before that, we shall review some building blocks of factor models in estimating the term structure of interest rates.

III.4 Factor models in estimating term structure of interest rates

It is an empirical question that asks how we summarize the information implied by yield curve at any point in time when faced with large amount of traded nominal bonds. In fact, from Figure III-8, we can see most movements in yields are due to level shifts and slope changes of the yield curve. This is in accordance with most finance literature on factor models claiming that only a small number of factors are sufficient to describe the systematic risks underling traded financial assets. Therefore, it is this empirical evidence that motivates us to model the yields dynamics for different maturities with some carefully constructed factors and their loadings (the coefficients of those factors). Besides these, using factor models also allow us to achieve the principle of parsimony in econometric modelling. Compared to the unrestricted Vector Autoregressive (VAR) models with large number of estimated coefficients, a properly constructed factor model with restrictions imposed on either factors or their loadings often provides better forecasting power.

Figure III-8: US 1 to 5 years Zero-Coupon-Yield level and slope (06/1952–12/2008)

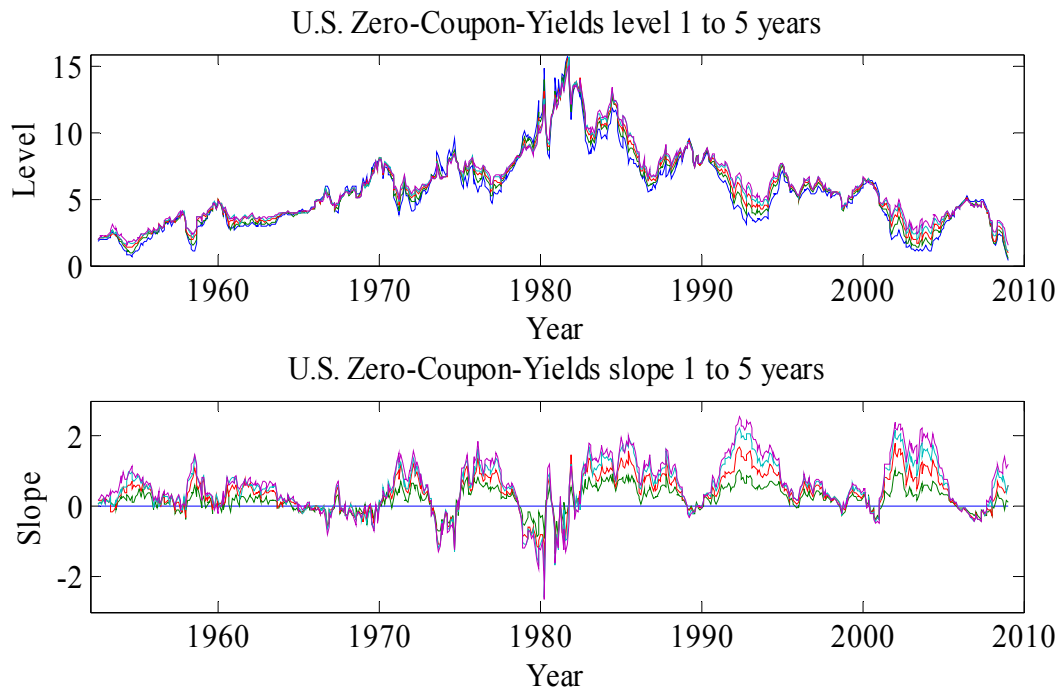


Figure III-9: UK 1 to 10 years Zero-Coupon-Yield level and slope (01/1970–09/2009)

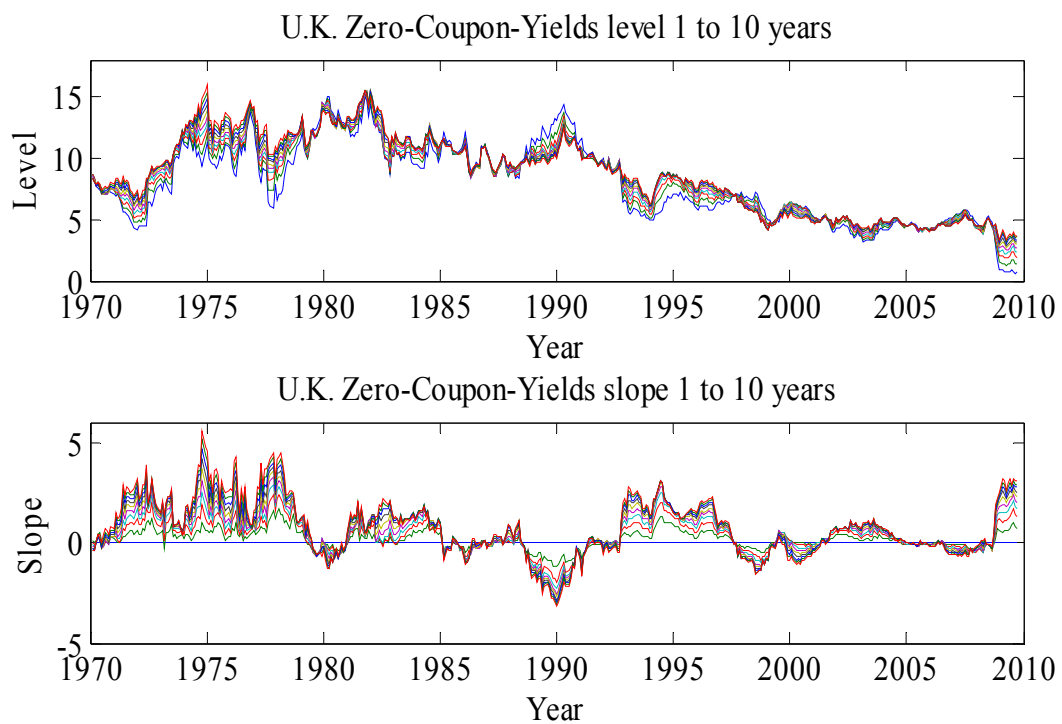
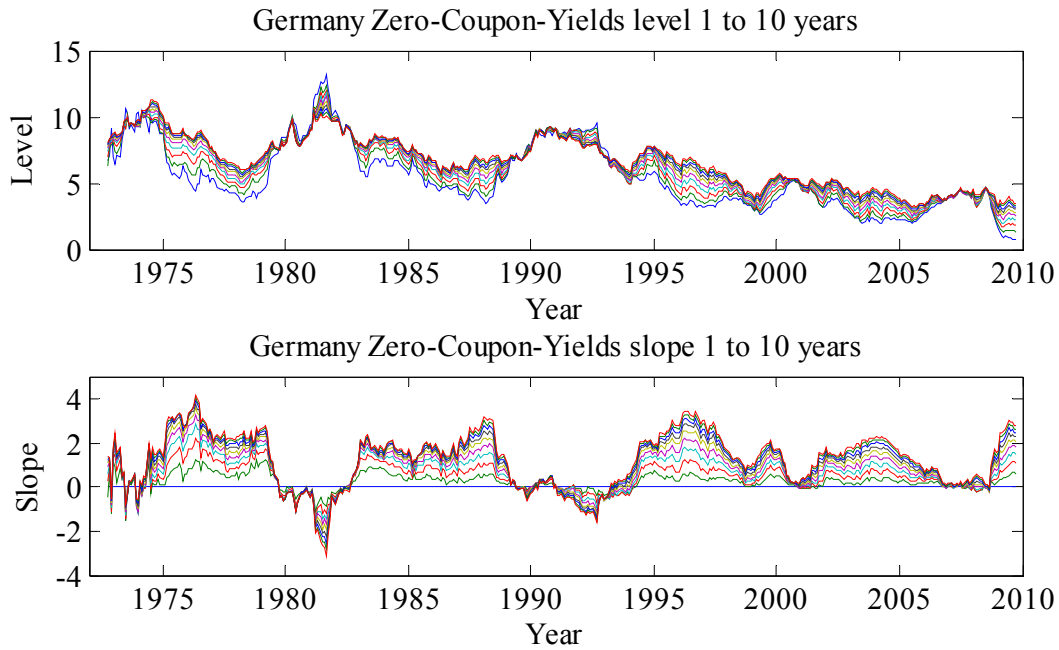


Figure III-10: German 1 to 10 years Zero-Coupon-Yield level and slope (09/1972–09/2009)



In constructing factor models, we notice that there are generally two approaches in the literature; either imposing restrictions on factors (like the principal component analysis) or restricting the factor loadings (like the Nelson-Siegel factor model). To start with constructing these models, it is worthy to discuss what the intuition behind the factor model. Consider an ad-hoc factor model which we want to use to describe the dynamics of yields in Figure III-8: if we define the level factor as the average of all yields and the slope factor as the difference between longest yield and shortest yield, we have the following equation for yields dynamics

$$y_{n,t} = c_n + l_n \times \text{level}_t + s_n \times \text{slope}_t + \varepsilon_{n,t} \quad (\text{III.21})$$

where c_n, l_n, s_n are the coefficients for level and slope factors, respectively. (III.21) states that, when all yields rise, the level factor becomes bigger and when the yield curve slopes upwards (downwards), the slope factor is positive (negative). Therefore, what this factor model emphasises is that there are just two portfolios of bond yields (equally weighted portfolio with all bonds and longing the longest while shorting the shortest) that matters, changes to other yields are just different linear combinations of the two portfolios. In other words, this is

close to the “complete market” hypothesis, which states that once we know what level_{*t*} and slope_{*t*} are, all other yields in the bonds market are replicable by using the two factors.

III.4.1 Principal components

The construction of the level and slope factors in the previous subsection is rather ad-hoc, and a more flexible way to construct these factors should have a mathematically sound background. Our problem is to construct factors that are not related to each other (that is the restriction we put on factors) to minimize the variance of errors, which is in contrast to the standard OLS regression methods where factors are fixed but coefficients are to be chosen to minimize the variance of error. Principal components analysis based on eigenvalue decomposition of matrices supplies us the tool to construct such mutually orthogonal factors. Consider $\Sigma_t = \text{cov}(y_{n,t}, y'_{n,t})$, which is the covariance of the yields across maturities; the eigenvalue decomposition of this n-by-n symmetric covariance matrix takes the form of

$$\Sigma_t = Q\Lambda Q' \quad (\text{III.22})$$

with Λ as the eigenvalues laying on diagonal and $Q'Q = QQ' = I$. Factors are constructed as $F_t = Q' y_{n,t}$, where the columns of Q could be interpreted as the weights assigned to each yield, or they can also be regarded as factor loadings since $y_{n,t} = Q \times F_t = Q \times Q' y_{n,t}$. The decomposition also implies that the factors (F_t) are uncorrelated with each other as

$$\text{cov}(F_t, F_t') = \text{cov}(Q' y_{n,t} y'_{n,t} Q) = Q' \text{cov}(y_{n,t} y'_{n,t}) Q = \Lambda \quad (\text{III.23})$$

and the diagonal elements of Λ are the variances for each factor.

Using the above decomposition, we can recover the level and slope factor in (III.21)

by restricting $Q' = \begin{bmatrix} \overbrace{1/n \quad \dots \quad 1/n}^n \\ -1 \quad 0 \dots 0 \quad 1 \end{bmatrix}$, which states that the weights assigned to each yield in

constructing the first factor are equal while the weights for longest yield in constructing the second factor is 1 and -1 for the shortest yield.

In reality, we can have n factors constructed using bond yields with maturities 1 to n, however, some of the factors are usually very small and can be ignored which leaves the movements of yields to be driven by just a few factors. Litterman and Scheinkman (1991)

found that over 98% of the variation in returns on government fixed income securities can be explained by three factors, which are labelled as level, slope and curvature. As the diagonal of Λ measures the variance of the factors, the percentage explained by the i -th factor is

$\lambda_i / \sum_{i=1}^n \lambda_i$ and this could be a good gauge on deciding whether to include more factors.

In Figure III-11 to Figure III-16, we plot the replicated US, UK and German yields using principal components. Evidently, as we include more principal components, the closer the replicated ones are to the observed ones. The corresponding replication residual squares, without a doubt, decreases when the number of principal components included in replication grows.

Figure III-11: Replicate the US zero coupon bond yields using principal components

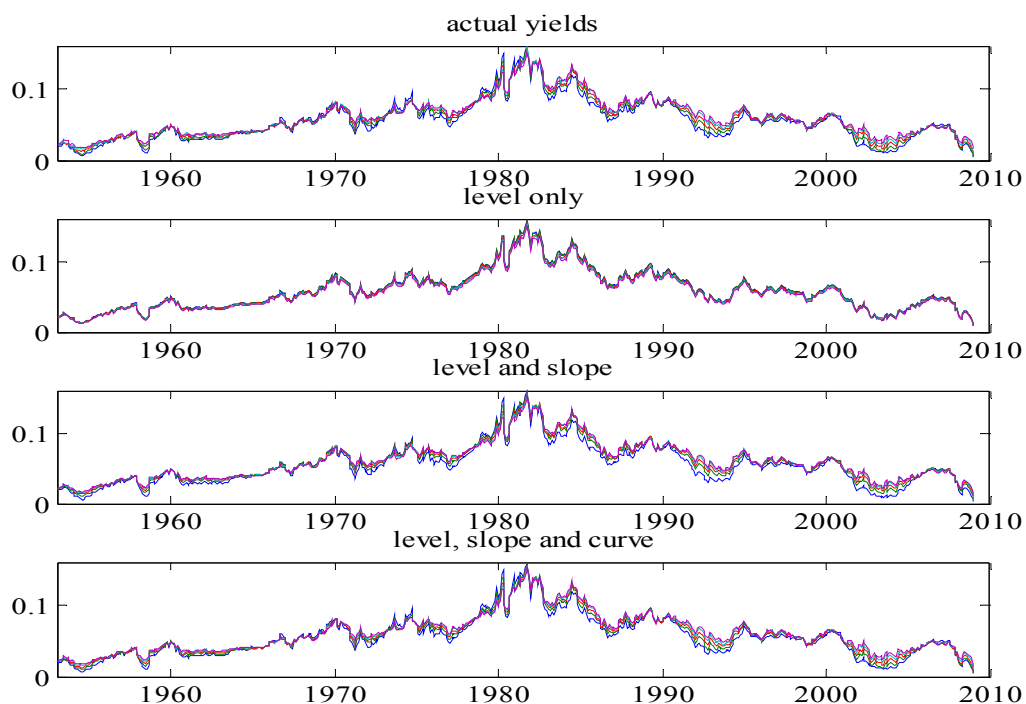


Figure III-12: US replication residuals square

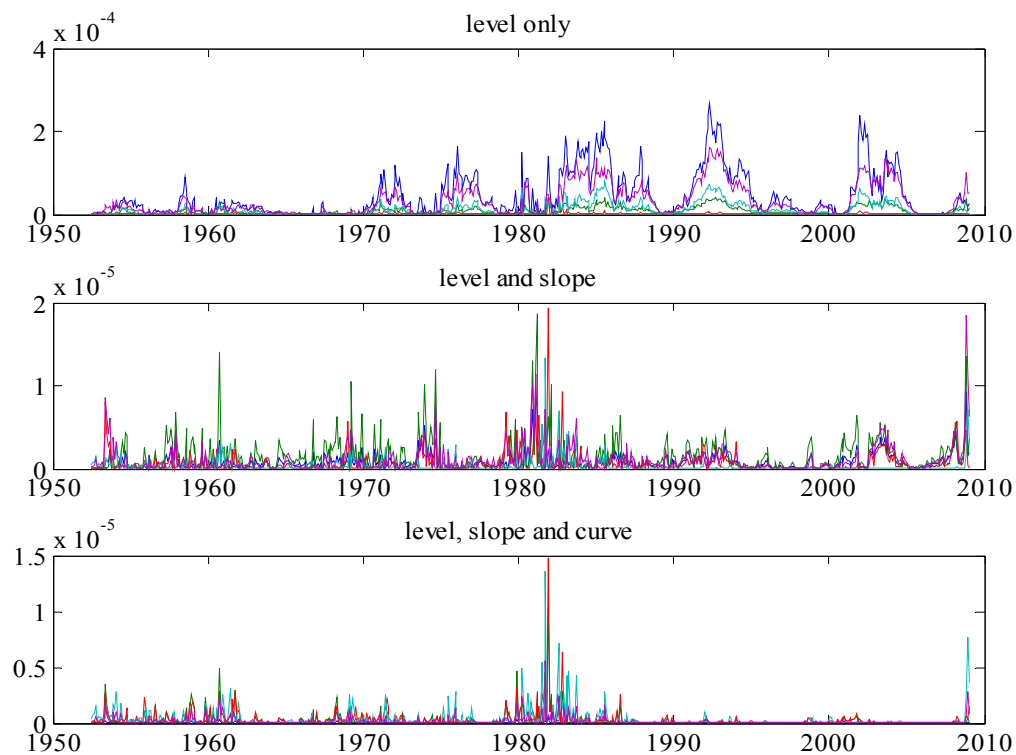


Figure III-13: Replicate the UK zero coupon bond yields using principal components

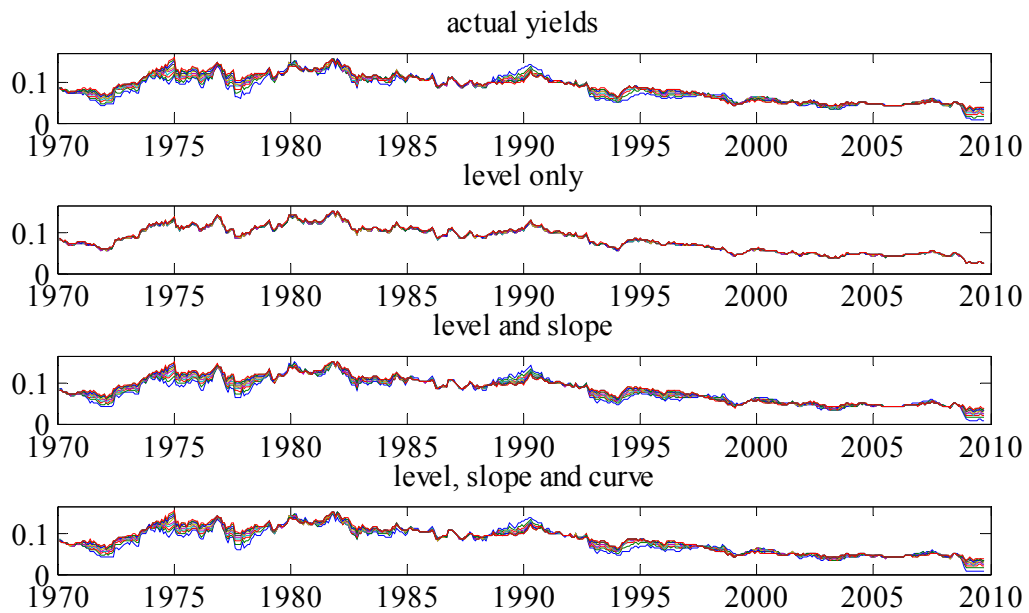


Figure III-14: UK replication residuals square

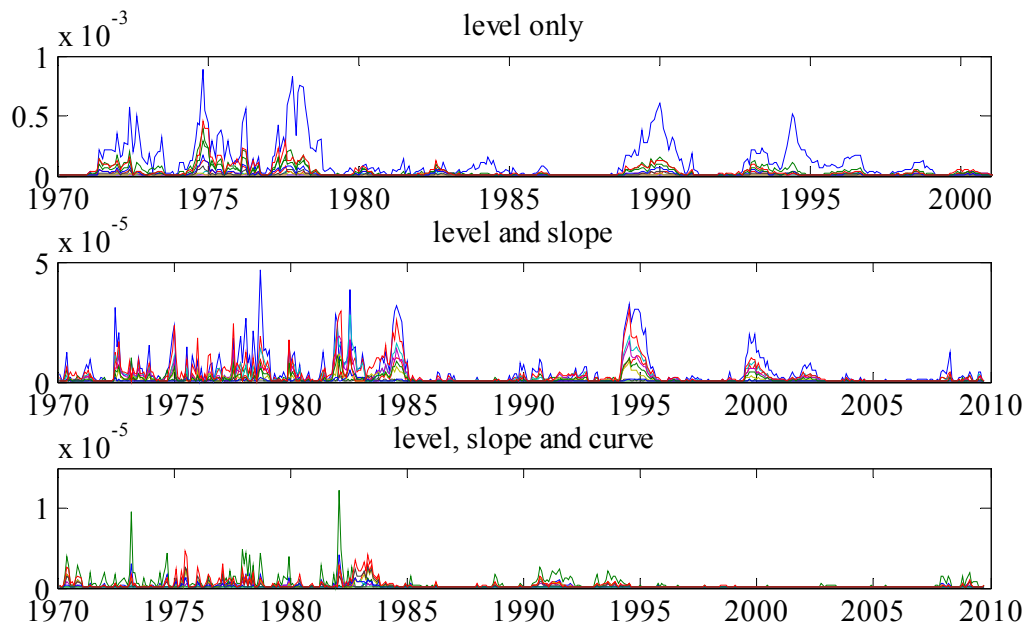


Figure III-15: Replicate the German zero coupon bond yields using principal components

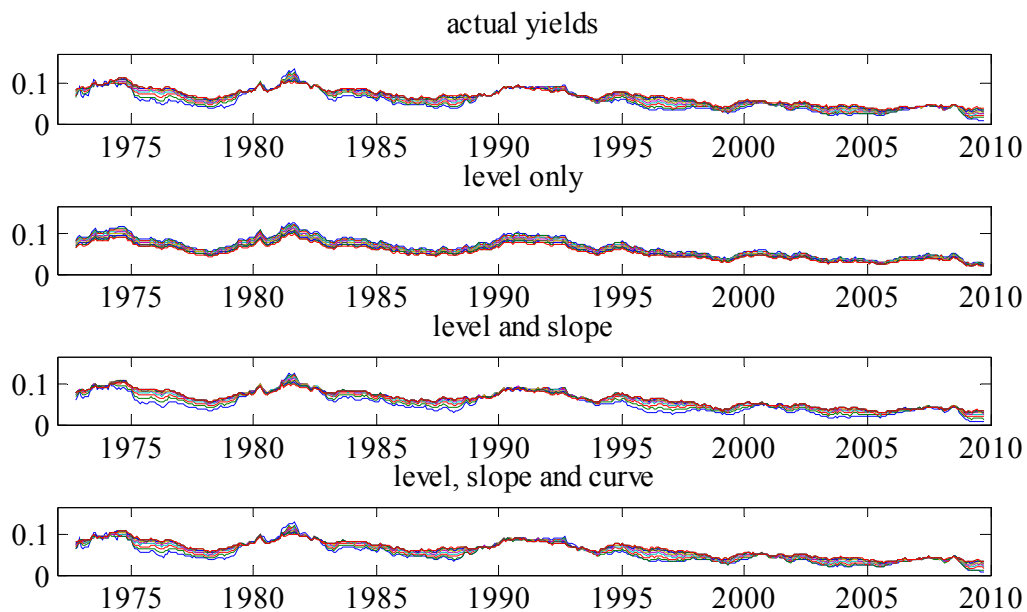
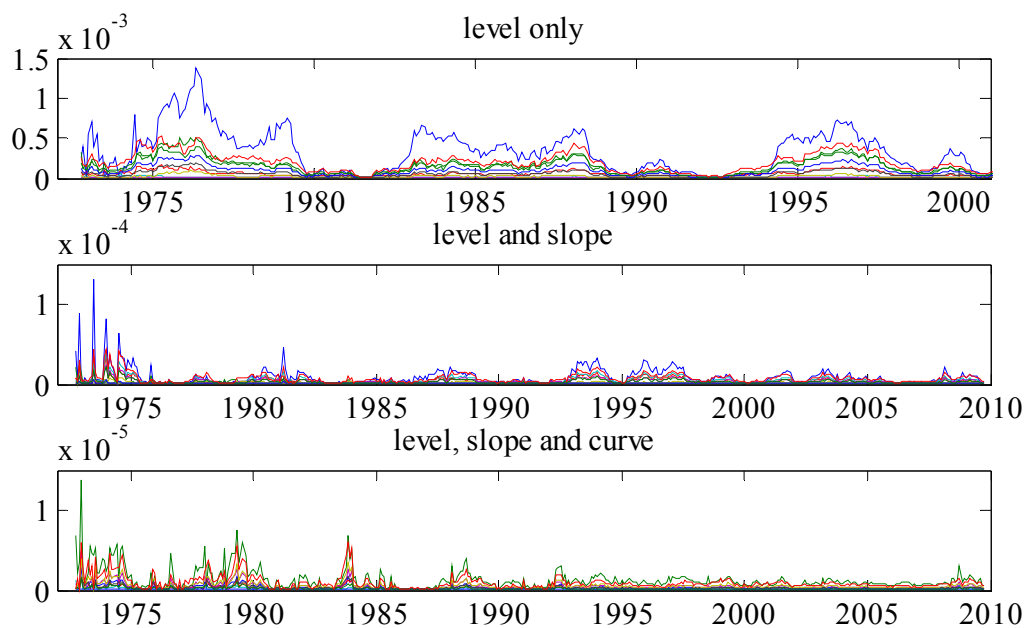


Figure III-16: German replication residuals square



III.4.2 Nelson-Siegel factor model

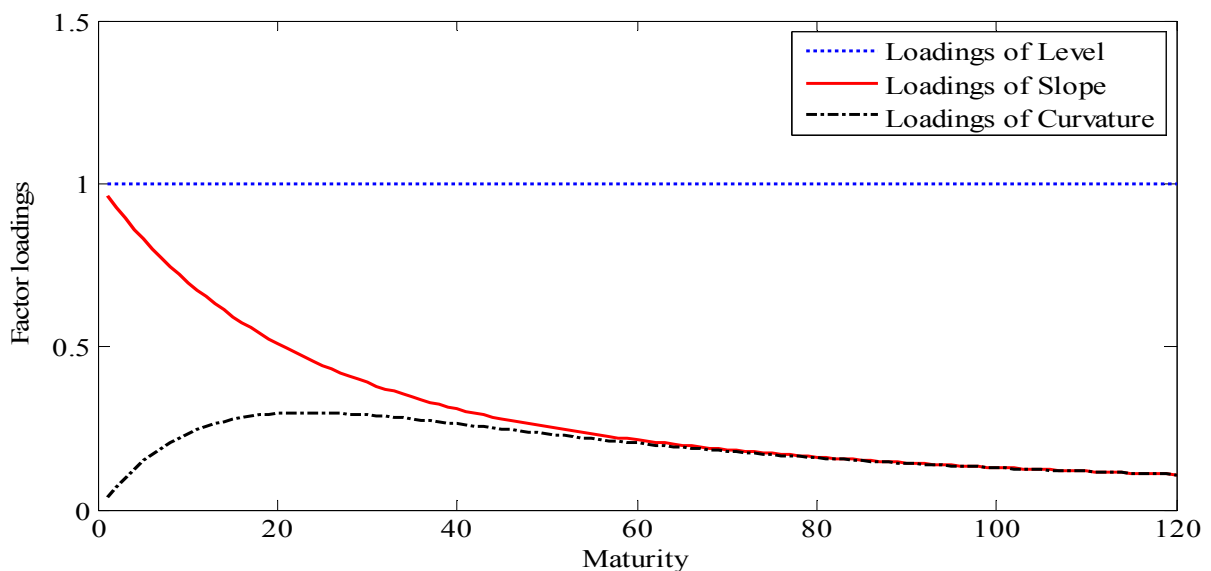
Although the principal component analysis in the previous subsection is flexible, it only allows us to analyse the bonds that we initially included in the sample. How do we interpolate bond yields with other maturities? Among many practitioners, Nelson and Siegel

(1987) factor model with variations is a popular choice. The dynamic Nelson-Siegel factor model takes the form

$$y_t^{(n)} = l_t + s_t \left(\frac{1 - \exp(-n/\lambda)}{-n/\lambda} \right) + c_t \left(\frac{1 - \exp(-n/\lambda)}{-n/\lambda} - \exp(-n/\lambda) \right) \quad (\text{III.24})$$

where n is the maturity of a bond; λ is a constant usually restricted at the median of the time series (but could be treated as an additional parameter to be estimated); l_t , s_t and c_t are time varying level, slope and curvature factors, and the terms that they are multiplied with are the corresponding factor loadings. In particular, l_t has a loading does not decay to zero in the limit, therefore it can be interpreted as a long-term factor (the level of yields); the loading on s_t decays monotonically towards zero from one, which can be interpreted as a short-term factor (the slope factor); c_t has a loading which first increases from zero and then decays to zero and hence is regarded as a medium-term factor that captures the curvature of the yield curve. Figure III-17 shows what the three factor loadings look like. Diebold and Li (2006) has shown that the curvature factor essentially captures well the difference between longest-middle yield spread and middle-shortest yield spread.

Figure III-17: What do Nelson-Siegel factor loadings look like



III.4.2.1 Estimation of the dynamic Nelson-Siegel factor model using Kalman filter

The conventional way of estimating (III.24) involves two steps. In the first step, the factor loadings are calculated, given the value of λ and maturity n . Then, the OLS estimation

on (III.24) is performed, for each time t , by treating the factor loadings as the regressors. Since the factor loadings obtained are constant over time and the yields are time varying, we obtain a different set of factor for each t . In step two, the factors are used to build up a VAR(1) process, from which the VAR coefficients are estimated.

Although the two-step procedure is convenient in getting a timely estimate and forecast, the uncertainties associated with factor loading calculation and factor extraction are not acknowledged in the second step. In contrast, an one-step maximum likelihood value estimation with Kalman filter is more efficient in terms of information utilisation. The simultaneous estimation of the measurement equation and state equation in a state-space form provides consistent and efficient parameters.

In this section, we demonstrate how to estimate a dynamic Nelson-Siegel factor model on the UK and German yield curves, using maximum likelihood estimation with Kalman filter. Let's rewrite (III.24) as

$$y_t^{(n)} = FL' \times NS_t + u_t^{(n)}, \quad (III.25)$$

where $FL' = \begin{bmatrix} 1 & \frac{1 - \exp(-n/\lambda)}{-n/\lambda} & \frac{1 - \exp(-n/\lambda)}{-n/\lambda} - \exp(-n/\lambda) \end{bmatrix}$, $NS_t = \begin{bmatrix} l_t & s_t & c_t \end{bmatrix}'$, and $u_t^{(n)}$

is assumed to be independently identically distributed with mean zero and variance Ω . Equation (III.25) is called the measurement equation, which links the observed yields to the unobserved factors through the measurement system matrices FL' . The dynamics of the unobserved factors, i.e. the Nelson-Siegel factors, is modelled in the state equation, which takes the form

$$NS_t = \alpha + \beta NS_{t-1} + e_t, \quad e_t \sim i.i.d.N(0, \Sigma). \quad (III.26)$$

In order to estimate equations (III.25) and (III.26) simultaneously, we need to use Kalman filter in the maximum likelihood estimation procedure. Kalman filter is a recursive procedure for calculating the optimal estimate of the unobserved variables, like the NS_t in (III.26), based on the appropriate information set assuming α , β and Σ are known *a priori*. The first step in a Kalman filter procedure is to form the optimal predictor of the next observation using prediction equations, given all the information currently available. In the second step a series of updating equations are used to incorporate the new observation into

estimates of the unobserved variables. It is well known that the Kalman filter provides an optimal solution to the problem of predicting and updating the unobserved variables in system like ours. If the observations ($y_t^{(n)}$ in our case) are assumed to be normally distributed and the current estimates of the unobserved variables are the best available, the predictor and the updated estimates will also be the best available. However, if the normality assumption does not hold, a similar result will hold, but only within the class of estimators and predictors that are linear in the observations.

Let's now derive the Kalman filter predicting and updating equations. First, we need to initialise the Kalman filter with carefully chosen initial values, that is to say the values of α , β , Ω and Σ at time $t-1$. For the factor loadings, which are functions of n and λ , if we treat λ as an extra parameter need to be estimated, we must also initialise it at time $t-1$. In this exercise, however, we set $\lambda = 0.077$ as in Diebold and Li (2006). The factor loadings, therefore, are constant in this case and can be used to back out the unobserved Nelson-Siegel factors. An OLS estimation of equation (III.26) is then followed, which leads us to a set of initial values for α , β , Ω and Σ at time $t-1$. To start the Kalman filter, we also need initial values for the Nelson-Siegel factors at time $t-1$. We set them to the unconditional mean of NS_t .

Given the initial values of α , β , Ω , Σ and $NS_{t-1|t-1}$ at time $t-1$, we can now derive the Kalman filter predicting equation as

$$NS_{t|t-1} = \alpha + \beta NS_{t-1|t-1}, \quad (\text{III.27})$$

where $NS_{t|t-1}$ is the predicted Nelson-Siegel factors at time t based on the information up to time $t-1$. The corresponding mean square error of the unobserved $NS_{t|t-1}$, $P_{t|t-1}$, is then calculated as

$$P_{t|t-1} = \beta NS_{t-1|t-1} \beta' + \Sigma. \quad (\text{III.28})$$

Denoting $\eta_{t|t-1}$ as the prediction error, we define it as

$$\eta_{t|t-1} = y_t^{(n)} - FL' \times NS_{t|t-1}. \quad (\text{III.29})$$

Subsequently, the conditional variance of the prediction error takes the form

$$f_{t|t-1} = FL' \times P_{t|t-1} \times FL + \Omega . \quad (\text{III.30})$$

Equations (III.27) to (III.30) constitute the prediction system of the Kalman filter. In order to update the unobserved factors, we write

$$NS_{t|t} = NS_{t|t-1} + \frac{P_{t|t-1} \times FL}{f_{t|t-1}} \eta_{t|t-1} , \quad (\text{III.31})$$

where $P_{t|t-1} \times FL / f_{t|t-1}$ is the Kalman gain, which determines the weight assigned to new information about the unobserved factors contained in the prediction error $\eta_{t|t-1}$. Finally, we can update the mean square error as

$$P_{t|t} = P_{t|t-1} - \frac{P_{t|t-1} \times FL \times FL' \times P_{t|t-1}}{f_{t|t-1}} . \quad (\text{III.32})$$

As a by-product of the predicting and updating process in the Kalman filter, we obtain the prediction error in (III.29) and its variance in (III.30), which can be used to calculate the log likelihood value as

$$l = -\frac{1}{2} \sum_{t=1}^T \ln(2\pi \times |f_{t|t-1}|) - \frac{1}{2} \sum_{t=1}^T \eta'_{t|t-1} f_{t|t-1}^{-1} \eta_{t|t-1} . \quad (\text{III.33})$$

The estimation of the state space system formed by (III.25) and (III.26), requires an optimization on a large set of parameters. In the measurement equation (III.25), since we fix λ to 0.077, the free parameters that are needed to be estimated have reduced to Ω , the covariance matrix of the added measurement errors. Like in Diebold and Li (2006), we assume the Ω matrix is diagonal, which implies that the variations of yields from different maturities are uncorrelated. This standard assumption also makes the estimation of the Ω matrix computationally simplified and tractable, given the large number of yields included in estimation. In the state equation (III.26), the 3-by-1 unconditional mean vector α contains 3 free parameters, and the 3-by-3 transition matrix β contains 9 free parameters. Since this VAR(1) system of the state equation describes the state dynamics which drives the stationary yields, the diagonal of the β matrix should be less than one in absolute values to keep the unobserved factors unexplosive. We, however, do not explicitly put these restrictions on the diagonal of the β matrix, but let the data tell us if these restrictions are preserved in data generating process. The 3-by-3 covariance matrix Σ is also unrestricted, which allows the

shocks to the three unobserved factors to be correlated contemporarily. There are 6 free parameters in matrix Σ , among which 3 are the variances for each of the three unobserved factors and the rest 3 parameters are covariance terms. In total, there are 28 free parameters in the state space system that are to be estimated. As told earlier, the initial values of these parameters are obtained from the two-step OLS estimation. Then, we maximise the log-likelihood function of (III.33) iteratively using the Nelder-Mead simplex algorithm as described in Lagarias, Reeds, Wright and Wright (1998).

III.4.2.2 Estimation results

We examine the UK and German treasury bond yields with maturities of 1 year to 10 years. The statistic properties of the two sets of data are discussed previously in section III.3.4. The estimated parameters with standard errors are shown in Table III-1.

For the UK market, the persistence of the level, slope and curvature factors, which are measured by the diagonal of the β matrix, are very high and highly significant, with estimated coefficients of 0.9963, 0.9406 and 0.8514, respectively. The cross-factor effects, on the other hand, seem less important, as a few cross-factor coefficients are insignificant. Yet, there is a small but significant negative effect of the level factor one-period before on the current curvature factor. In contrast with the decreasing persistence as move along the diagonal of β matrix from level to slope to curvature, the estimated variances of the three factors show an increasing pattern along the diagonal of the Σ matrix. In addition, the cross-factor transition shocks are all negative as indicated by the negative covariance terms.

For the German market, which is similar to the UK market, the three factors also show a decreased persistence along the diagonal of the β matrix from level to slope to curvature, and an increasing variance along the diagonal of Σ matrix. What is different in the German market is that the level factor one-period before tends to affect the current slope factor negatively. In addition, the contemporary covariance between the level and slope factors is positive.

Table III-1: UK and German Nelson-Siegel model estimates and standard errors

	UK	German
α_1	4.091E-05 (5.711E-05)	5.699E-05 (6.056E-05)
α_2	-1.43E-04 (7.800E-05)	2.928E-04 (1.275E-04)
α_3	1.510E-04 (2.134E-04)	-5.811E-04 (2.768E-04)
$\begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{bmatrix}$	$\begin{bmatrix} 9.963E-01 & 1.688E-04 & 1.550E-02 \\ (7.649E-03) & (1.021E-02) & (6.569E-03) \\ 1.801E-02 & 9.406E-01 & 3.802E-02 \\ (1.048E-02) & (1.424E-02) & (9.177E-03) \\ -4.615E-02 & 4.251E-03 & 8.514E-01 \\ (2.868E-02) & (3.910E-02) & (2.513E-02) \end{bmatrix}$	$\begin{bmatrix} 9.929E-01 & 5.523E-03 & 9.245E-03 \\ (9.339E-03) & (7.362E-03) & (4.717E-03) \\ -4.992E-02 & 8.898E-01 & 6.829E-02 \\ (1.993E-02) & (1.809E-02) & (1.129E-02) \\ 7.842E-02 & 1.312E-01 & 8.651E-01 \\ (4.321E-02) & (3.879E-02) & (2.424E-02) \end{bmatrix}$
$\begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \Sigma_{1,3} \\ \Sigma_{2,1} & \Sigma_{2,2} & \Sigma_{2,3} \\ \Sigma_{3,1} & \Sigma_{3,2} & \Sigma_{3,3} \end{bmatrix}$	$\begin{bmatrix} 1.642E-07 & -4.829E-08 & -3.003E-07 \\ (2.381E-06) & (1.057E-08) & (1.057E-08) \\ -4.829E-08 & 3.175E-07 & -4.174E-08 \\ (1.057E-08) & (2.559E-06) & (1.057E-08) \\ -3.003E-07 & -4.174E-08 & 2.406E-06 \\ (1.057E-08) & (1.057E-08) & (3.138E-06) \end{bmatrix}$	$\begin{bmatrix} 5.786E-08 & 2.862E-08 & -1.591E-07 \\ (2.685E-06) & (1.057E-08) & (1.057E-08) \\ 2.862E-08 & 3.468E-07 & -4.775E-07 \\ (1.057E-08) & (2.616E-06) & (1.057E-08) \\ -1.591E-07 & -4.775E-07 & 1.602E-06 \\ (1.057E-08) & (1.057E-08) & (2.898E-06) \end{bmatrix}$
$\Omega^{(1)}$	1.911E-09 (2.384E-06)	9.067E-10 (2.639E-06)
$\Omega^{(2)}$	1.620E-08 (2.002E-06)	8.499E-09 (1.989E-06)
$\Omega^{(3)}$	4.835E-09 (1.871E-06)	2.017E-09 (1.859E-06)
$\Omega^{(4)}$	5.236E-13 (1.681E-06)	5.775E-18 (1.675E-06)
$\Omega^{(5)}$	1.826E-09 (1.761E-06)	6.946E-10 (1.742E-06)
$\Omega^{(6)}$	2.621E-09 (1.796E-06)	9.607E-10 (1.769E-06)
$\Omega^{(7)}$	1.280E-09 (1.730E-06)	4.183E-10 (1.704E-06)
$\Omega^{(8)}$	2.193E-20 (1.649E-06)	2.228E-22 (1.612E-06)
$\Omega^{(9)}$	2.380E-09 (1.798E-06)	7.210E-10 (1.750E-06)
$\Omega^{(10)}$	1.215E-08 (1.957E-06)	3.262E-09 (1.877E-06)
Log Likelihood	37323.952	36863.646

Note: Standard errors which are calculated using finite-difference method are shown in parentheses.

The superscripts of Ω indicate maturities in years.

Using the Kalman filter, we are able to extract the unobserved level, slope and curvature factors from the observed yields. In Figure III-18 and Figure III-19, we plot the extracted factors for the UK and German markets, respectively. The level factors in both markets are positive and highly persistent (and less volatile) in comparison to the slope and

curvature factors. Furthermore, the level factors show decreasing trends over time in both markets. In contrast, the slope and curvature factors in both markets show less persistence but more volatile fluctuations, assuming both positive and negative values.

The remaining estimates in Table III-1 are the variances of the measurement errors in (III.25). None of them are statistically significant, as the values are very small. Since we assume constant variance of those measurement errors in both markets, we plot the residual squares of the fitted UK and German yields in Figure III-20 and Figure III-21 to see if the constant variance assumption is too simplified. Clearly, there seems to be a high volatility period followed by a low volatility period in both markets, although the high volatility period in the German market tends to be a temporary phenomenon. Despite the weakness of the constant variance assumption, our Nelson-Siegel factor model fits the curve remarkably well, as evidently shown in Figure III-22 and Figure III-23.

Figure III-18: UK Nelson-Siegel factors

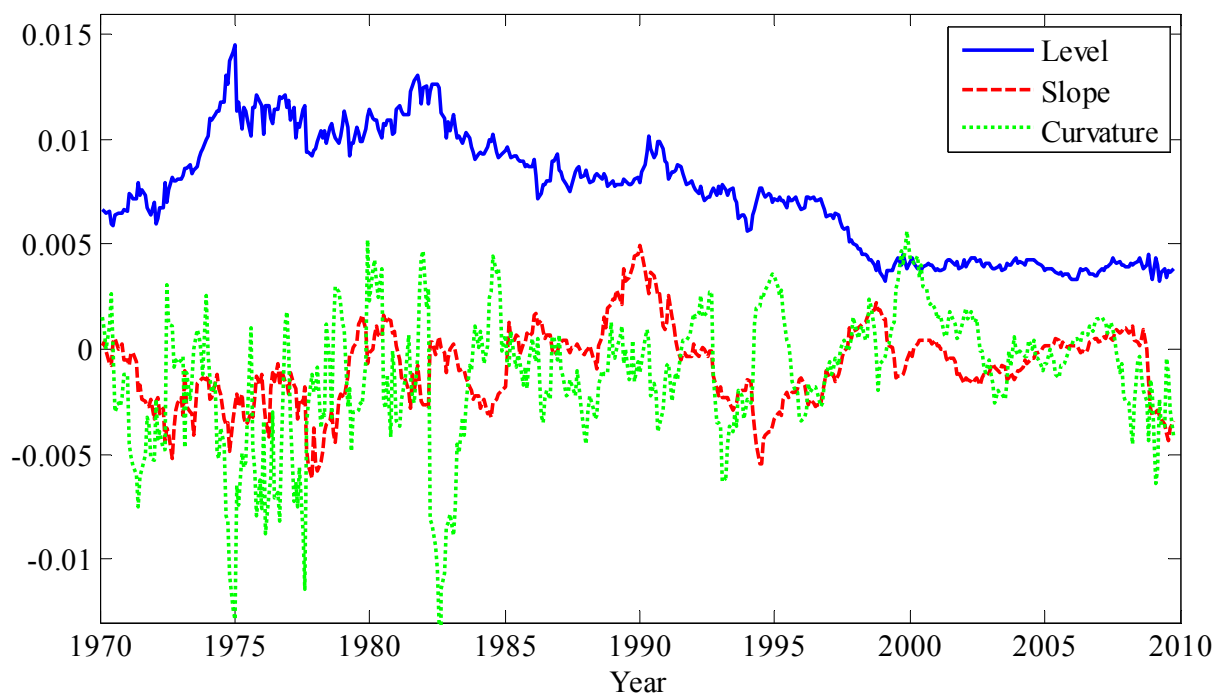


Figure III-19: German Nelson-Siegel factors

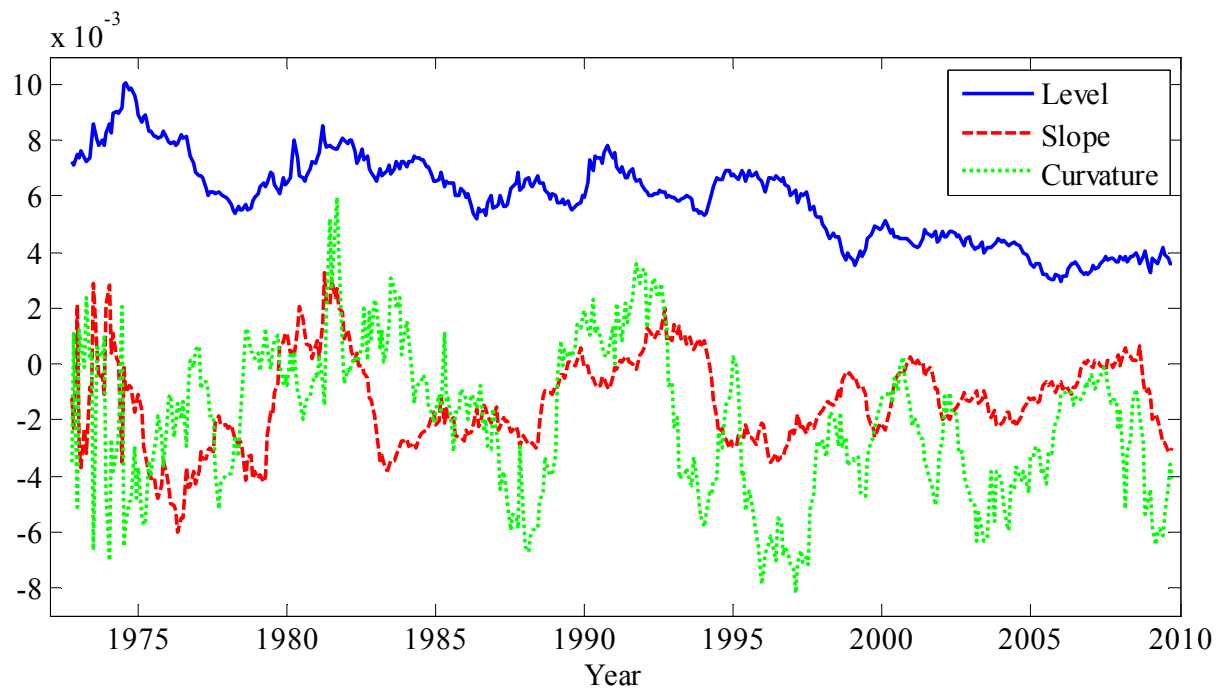


Figure III-20: Residual square of the fitted UK yields

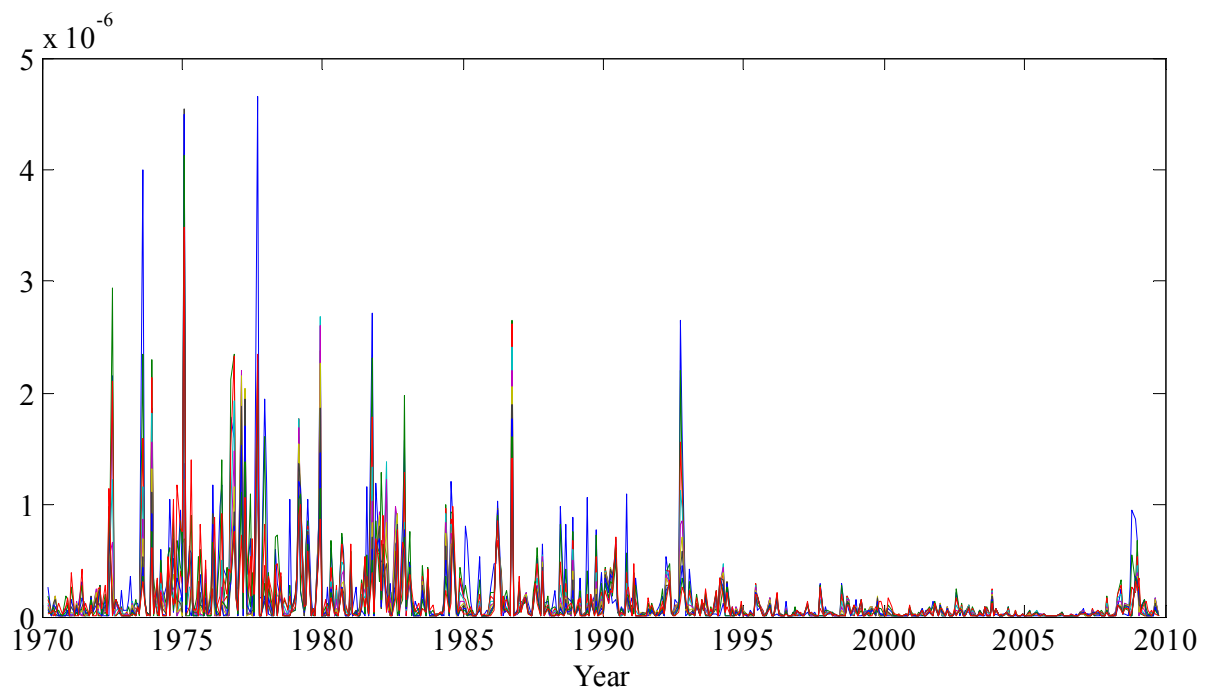


Figure III-21: Residual square of the fitted German yields

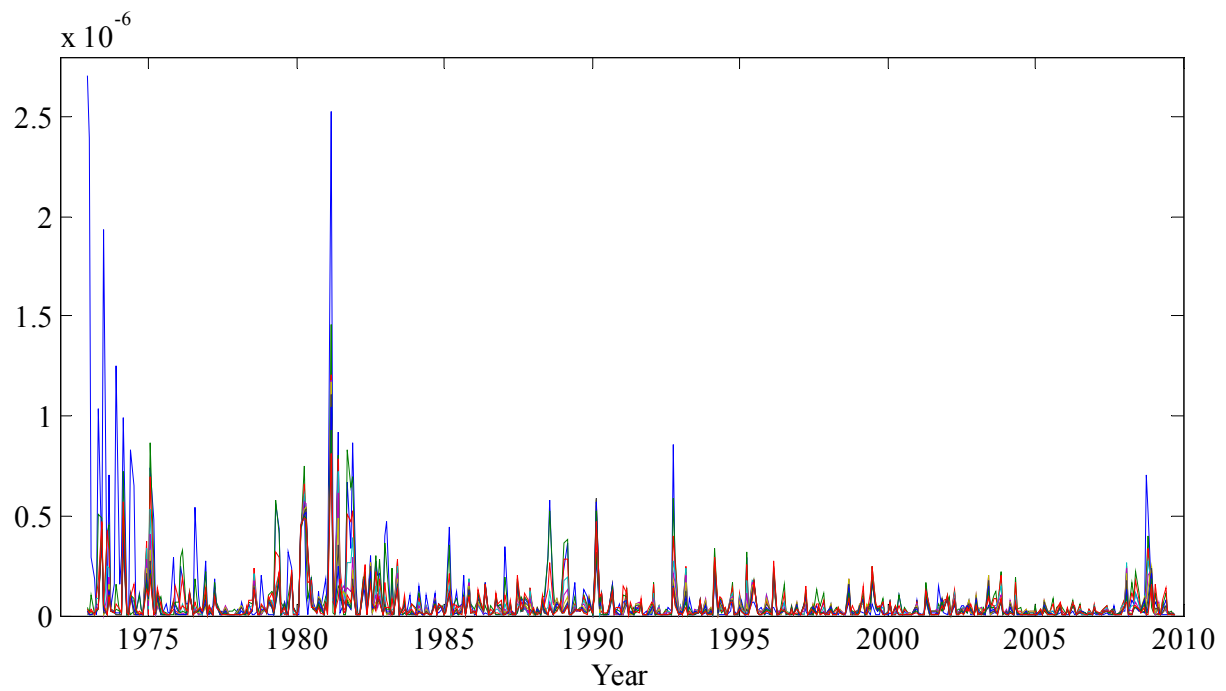


Figure III-22: Observed and fitted UK yields

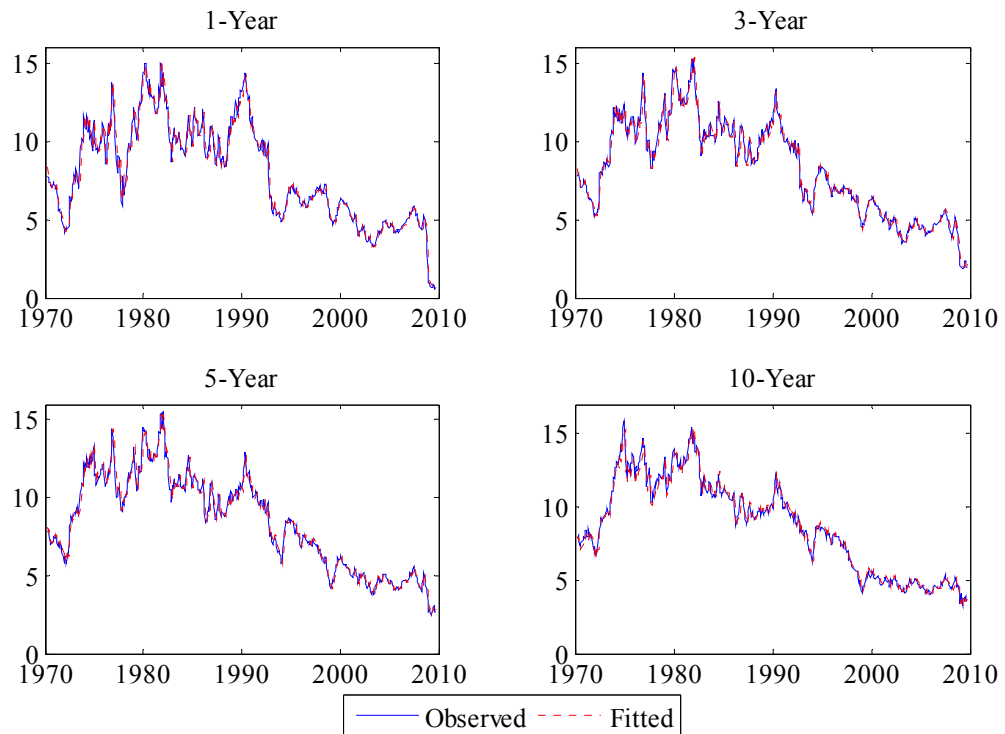
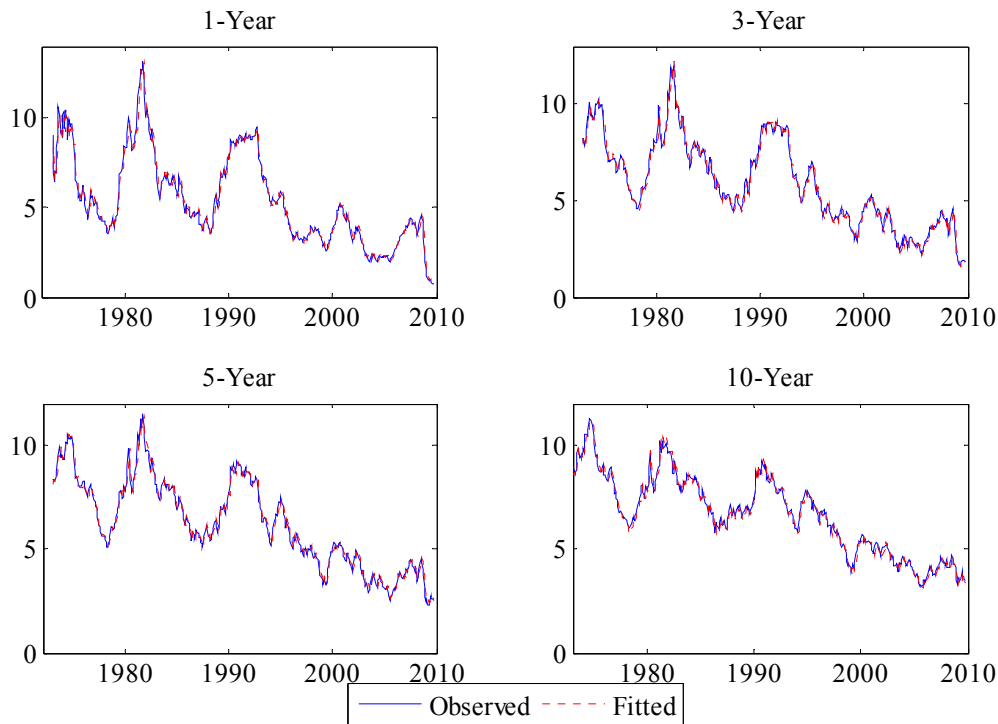


Figure III-23: Observed and fitted German yields



III.5 No-arbitrage factor models

The principal component analysis and Nelson-Siegel factor models are general and flexible to accommodate different shapes of the yield curve. Nevertheless, they do not put no-arbitrage restrictions across bond maturities, and they do not allow explicitly modelling the term premia. In fact, the standard response in empirical finance literature to the rejection of expectation hypothesis has been modelling the term structure based on the no-arbitrage assumption across different maturities of bonds. In this section, we introduce the equivalence between no-arbitrage conditions and the existence of a strictly positive pricing kernel. Unlike the majorities modelling the term structure in a continuous-time framework, we use an econometrical discrete-time approach. We compare different specifications of the no-arbitrage affine class models, with special attention paid on the treatment to time-varying term premia. We discuss a simple one-factor model and extend that to include regime switching of the short rate in Chapter IV. By modelling regime switches of the short rate, we would generate non-normal distribution that is consistent with evidence on nonlinearities in conditional moments. In general, the Markov-switching models are outside of the affine class, however, for some special cases, affine solutions can still be obtained (Dai and Singleton (2003), Singleton (2006) and Ang, Bekaert and Wei (2008a)).

III.5.1 No-arbitrage conditions and the stochastic discount factor in term structure model

There are many definitions of the opportunities of arbitrage in the literature, which are essentially equivalent, but are interpreted differently. The basic definition of arbitrage refers to a situation where ‘free-lunch’ cannot be eliminated, or it is possible to have a dynamic portfolio that does not cost anything but definitely offers some positive payoffs with a positive probability. To proceed to a precise definition of the arbitrage opportunities, we need to introduce some notions.

Let's express the uncertainty involved in any investment under the discrete-time framework as a probability space (Ω, \mathcal{F}, P) , where Ω is the sample space consists of random events which may potentially happened; the σ -algebra \mathcal{F} , also called the σ -field, is a collection of events (subsets of Ω); the probability measure P is a function of \mathcal{F} , such that $P: \mathcal{F} \rightarrow [0, 1]$, tells us the probability of each event. Having the events indexed by $0, 1, \dots, T$, the sub-sigma algebra $\mathcal{F}_t \subseteq \mathcal{F}$ represents the information available at time t . Accordingly, the evolution of information over time follows the information filtration $\mathcal{F} = \{\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_T\}$, with $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$. If we denote P_t (the price of a security) as a sequence of random vectors, such that $P_t = \{P_0, P_1, \dots, P_T\}$, this process is said to be adapted to the information filtration \mathcal{F} if each component of P_t is a random variable with respect to (Ω, \mathcal{F}_t) . For N securities in a portfolio, the prices of those N securities at time t can be collected in the vector $P_t = (P_t^{(1)}, \dots, P_t^{(N)})'$. The trading strategy $H = (H_{1,t}, \dots, H_{N,t})'$ is the amount of securities held for each corresponding securities, where $H_{i,t}$ represents the amount of security i held in a portfolio within time period $t < s \leq t+1$, that is, positions of those securities are built after observing the information at time t and are held until the end of $t+1$. Position on each security can be positive or negative, where the later case refers to a short sell of the security. The changes of each security's positions in a portfolio is denoted by $\omega = \{\omega_t\}$ such that $\omega_t = H_t - H_{t-1}$. Closely related to the changes in positions of securities in a portfolio is the holding-period gain process $\delta = \{\delta_t\}$ that is defined as $\delta_t = (H_{t-1} - H_t)' P_t$. If δ_t is positive, the trading strategy over time period $(t-1, t]$ results a withdrawal of gain, if δ_t is negative, the trading strategy results an injection of funds to finance the new portfolio. At time $t = 0$, if

we assume $H_{t-1} = 0$, we have $\delta_0 = (0 - H_0)' P_0$ as the initial fund to finance the portfolio, while at time $t = T$, if we assume $H_T = 0$, the final value of the portfolio is given by $\delta_T = (H_{T-1} - 0)' P_T$. A trading strategy with $\delta_t = 0$ for $t = 1, \dots, T-1$, that is, no exogenous infusion or withdrawal of holding period gains from the portfolio, is called self-financing strategy, i.e. the purchasing of new portfolios must be financed by selling the old ones. Finally, a claim process, $\varsigma = \{\varsigma_t\}$, determines the entitled payoff to the owner of such a claim for $t = 1, \dots, T$. If this claim is replicable or hedgeable, there must be a trading strategy H such that $\delta_t(H) = \varsigma_t$ for each t , and the market with N securities is complete if every claim can be replicated.

Now, a formal definition of arbitrage strategy considers a trading strategy if $\delta_t(H) \geq 0$ for $t = 0, 1, \dots, T$ and $P(\delta_t(H) > 0) > 0$ for at least one t in $\{0, 1, \dots, T\}$. There exists no arbitrage strategy in a market if for each $t = 1, \dots, T$ and each $i = 1, \dots, N$ securities, there exists an F_t measurable stochastic discount factor M_t with $P(M_t > 0) = 1$ such that $P_{t-1}^{(i)} = E_{t-1}(M_t P_t^{(i)})$. The stochastic discount factor M_t , which is also called the pricing kernel, can be explicitly modelled as the marginal substitution rate at which investors are willing to substitute consumption at later date for consumption at current date, such as $M_t = \beta \left(\frac{u'(C_t)}{u'(C_{t-1})} \right)$. C_t is the consumption at date t , β is the time discount factor. The dynamics of the pricing kernel need a special treatment as M_t cannot be negative as we know $(u'(C_t)/u'(C_{t-1}))$ must be positive (assuming monotonic utility function) to keep out arbitrage opportunities.

To see how this economic interpretation of marginal substitution rate of consumption is linked to the fundamental pricing equation

$$P_{t-1}^{(i)} = E_{t-1}(M_t P_t^{(i)}), \quad (\text{III.34})$$

consider the utility function of an agent's preference over a consumption stream $C = \{C_t\}$ as

$$U(C) = \sum_{t=0}^T \beta^t E_0(u(C_t)). \quad (\text{III.35})$$

Assuming the agent receives exogenous income e_t at time t and chooses to either consume it or invest in portfolios. The resulting budget constraint can be written as

$$C_t + P_t H'_t = e_t + P_t H'_{t-1}, \quad (\text{III.36})$$

where the left-hand-side of the equation represents the consumption and investment decision at time t and the right-hand-side is the available funds to the agent consisting of the exogenous income and gain from last period's investment. The optimal allocation to consumption and investment $\{C^*, H^*\}$, can be found by solving the following maximization problem

$$\begin{aligned} \max_{C, H} U(C), \\ \text{s.t. } C_t = e_t - P_t (H'_t - H'_{t-1}), \\ C_t \geq 0, \quad t = 0, 1, \dots, T. \end{aligned} \quad (\text{III.37})$$

The solution to (III.37) is obtained from the first order condition which is

$$\begin{aligned} P_{t-1}^{(i)} = E_{t-1} \left(\beta \frac{u'(C_t)}{u'(C_{t-1})} P_t^{(i)} \right), \\ \text{for } i = 1, \dots, N. \end{aligned} \quad (\text{III.38})$$

It is clear if we compare (III.38) with the fundamental pricing equation that the stochastic discount factor $M_t = \beta \left(\frac{u'(C_t)}{u'(C_{t-1})} \right)$.

The stochastic discount factor, although is required to be positive to keep out arbitrage opportunities, it is however not required to be unique. It is unique only if the market is complete.

Given a positive and unique stochastic discount factor, the majority of asset pricing problem can be solved using the fundamental pricing equation. In the light of the consumption based explanation according to Cochrane (2005), any specification of the stochastic discount factor can be interpreted as a proxy for marginal utility.

In term structure of interest rate models, we will use stochastic discount factor approach to price a zero coupon bond at time t that has n periods left until maturity, denoted as $P_t^{(n)}$. In the next period, this bond has a price of $P_{t+1}^{(n-1)}$ with $n-1$ periods left until

maturity. The intertemporal pricing relationship between the two prices, according to (III.34), can be written as

$$P_t^{(n+1)} = E_t \left(M_{t+1} P_{t+1}^{(n)} \right). \quad (\text{III.39})$$

The price of a bond which matures in the next period (assuming the zero coupon bond pays off one unit at maturity, $P_{t+1}^{(0)} = 1$) is directly priced as $P_t^{(1)} = E_t (M_{t+1})$, and with repeated substitution, the price of a bond with two periods maturity can be written as

$$\begin{aligned} P_t^{(2)} &= E_t \left(M_{t+1} P_{t+1}^{(1)} \right) \\ &= E_t \left(M_{t+1} E_{t+1} (M_{t+2}) \right) \\ &= E_t \left(E_{t+1} (M_{t+1} M_{t+2}) \right) \\ &= E_t (M_{t+1} M_{t+2}), \end{aligned}$$

where the last equality follows from the law of iterated expectations. Analogously, a bond with n periods left until maturity can be priced as the time t conditional expectation of the product of each period's stochastic discount factor until the maturity:

$$P_t^{(n)} = E_t (M_{t+1} M_{t+2} \cdots M_{t+n}). \quad (\text{III.40})$$

To guarantee the positivity of the stochastic discount factor, it is suffice to take natural logarithm of it and model $\ln M_{t+1}$ instead:

$$\ln M_{t+1} = f(X_t) + e_{t+1}.$$

The specification of the conditional mean of $\ln M_{t+1}$ depends on the evolution of some underpinning state vector X_t and unexpected shock e_t . A trial solution to (III.40) may be proposed as taking the following functional relationship

$$P_t^{(n)} = f(X_t, t, \theta(n)) \quad (\text{III.41})$$

where $\theta(n)$ consists of maturity dependent parameters. If the choice of $\theta(n)$ guarantees (III.39) to be hold as an identity for all t and n , (III.41) as a guess of solution function to the fundamental pricing equation is succeed.

The conventional no-arbitrage affine term structure model consists of three main components: a transition equation that describes the dynamics of latent state vector X_t to be related to the prices of bonds; an equation defining the relationship between one-period short rate and the latent factors; and a pricing kernel equation relating the term premia with shocks to latent factors. For such a system of equations, the yield of a nominal bond is affine to these latent factors. The simplest no-arbitrage affine term structure model is the so-called short rate model, which we consider next.

III.5.2 Short rate model

Short rate models assume that the term structure of interest rates is driven by one factor only – the short rate. With a direct specification on one-to-one correspondence between the short rate and the single factor, these models are mainly motivated by the tractability of the models and the widely observed features of the short rate process like mean-reversion, non-negativity and the dependence of the volatility of short rate on its level. Two very popular, yet, very simple specifications on the short rate dynamics are known as the Vasicek (1977) model and Cox, Ingersoll and Ross (1985) model (hereafter CIR).

In Vasicek (1977) specification, the short rate is driven by a mean-reversion factor - x_{t+1} - with constant volatility, which in discrete-time formation can be written as

$$x_{t+1} - x_t = k(\delta - x_t) + \sigma \varepsilon_{t+1} \quad (\text{III.42})$$

The pricing kernel that allows us to introduce risk premia is modelled as

$$M_{t+1} = \exp\left(-x_t - \frac{1}{2}\lambda^2\sigma^2 - \lambda\sigma\varepsilon_{t+1}\right) \quad (\text{III.43})$$

The intuition is that under no-arbitrage condition, asset prices are generated as in $P = E(Mx)$, where P is asset price, x is the final pay-off of the asset. To link the short rate with the pricing kernel, in (III.42), we let the short rate to shift the mean of $\ln M_{t+1}$, which gives a slow change of M over time. The term $-\frac{1}{2}\lambda^2\sigma^2$ captures the mean adjustment of $\ln M_{t+1}$ which we will see later that it is built-in on purpose to reveal the fact that the single factor is the short rate. Pricing kernel is also known as stochastic discount factor, where the term “stochastic” comes from the fact in (III.42) that the external shock to the short rate equation,

$\sigma\mathcal{E}_{t+1}$, is introduced to the pricing kernel equation with its market price of risk λ . ($\sigma\mathcal{E}_{t+1}$ is assumed to be normally distributed i.i.d. variable, λ is assumed to be constant.)

The Vasicek (1977) model is the simplest affine model of the term structure of interest rate. By “affine”, it means that the bond yield is a linear function of the underlying state factors. In Vasicek (1977), the underlying state factor is the short rate, which can be seen from the one-period bond pricing equation that

$$\begin{aligned}
 y_{1,t} &= -\ln P_t^{(1)} = -\ln(E_t(M_{t+1} \times 1)) = -\ln(E_t(\exp(\ln M_{t+1}))) \\
 &= -\ln\left(\exp\left(E_t(\ln M_{t+1}) + \frac{1}{2}\text{var}(\ln M_{t+1})\right)\right) \\
 &= -\ln\left(\exp\left(\left(-\frac{1}{2}\lambda^2\sigma^2 - x_t\right) + \left(\frac{1}{2}\lambda^2\sigma^2\right)\right)\right) \\
 &= -\ln(\exp(-x_t)) = x_t
 \end{aligned} \tag{III.44}$$

where the first line reveals the market fact that the maturity payoff of the one-period bond is 1, the second line shows the mathematical fact that $E(\exp(x)) = \exp(\mu + \sigma^2/2)$ for $x \sim N(\mu, \sigma^2)$, and the third line is derived from (III.42) above. In a similar way, the yield for a bond with maturity of multiple time periods can be derived recursively using (III.44). Although the above model is simple, it reveals some basic facts about the dynamics of the term structure of interest rates. First, it allows all other rates to move together when the short rate moves, while maintaining the fact that longer ones move less than the shorter ones. Second, it explicitly models the term premia in the specification of the pricing kernel, though it is kept as a constant. In order to make the term premia time varying, we need the variance of the pricing kernel to vary over time. A straightforward extension is to let the market price of risk vary over time. As an example, a time-varying market price of risk would be read as $\lambda_t = \lambda_1 + \lambda_2 x_t$, that shows if the short rate rises, then λ_t rises, which leads a larger variance of the pricing kernel.

Apart from the extension to allow a time-varying term premia, Vasicek (1977) can also be extended in several other ways. One variation is to include the time-varying volatility, as we often see in financial markets that some financial variables (like the interest rates) are sometimes more, and sometimes less volatile. The Cox, *et al.* (1985) model is such an extension that it allows the conditional variance of the short rate to depend on the square root

of itself, which is also appealing for that it excludes the possibility of observing negative short rates. Analogous to (III.42), CIR defines the dynamics of the single factor that drives the short rate and pricing kernel as

$$\begin{aligned} x_{t+1} - x_t &= k(\delta - x_t) + \sigma\sqrt{x_t}\varepsilon_{t+1} \\ M_{t+1} &= \exp\left(-r_t - \frac{x_t}{2}\left(\frac{\lambda}{\sigma}\right)^2 - \left(\frac{\lambda}{\sigma}\right)\sqrt{x_t}\varepsilon_{t+1}\right) \end{aligned} \quad (\text{III.45})$$

Although the positive short rate claim in CIR model is plausible, relating the short rate's volatility sensitivity to its level by square-root is somehow restrictive. A more general approach would let the sensitivity parameter be a free parameter to be estimated as in $x_{t+1} - x_t = k(\delta - x_t) + \sigma x_t^\gamma \varepsilon_{t+1}$. The sensitivity parameter estimated in Chan, Karolyi, Longstaff and Sanders (1992) (hereafter CKLS) was 1.5, which indicates an explosive and persistent behaviour of short rate volatility. Many studies argue that this explosive parameter is due to the structural changes happened in 1974 (OPEC oil shocks), 1979 – 1982 (monetary experiment of Federal Reserve) and 1987 (stock market crash), which would significantly bias the estimate of the sensitivity parameter. Therefore, alternative models for volatility specification in short rate models are proposed in several directions. In Chapter IV, we shall discuss the Markov-switching extension that is generally believed as a powerful tool to tackle the structural break and nonlinearity in short rate dynamics.

III.5.3 Multi-factor affine model

The main advantage of the short-rate models is their simplicity that the entire yield curve is modelled as a function of just one state variable – the short rate as a proxy. However, there are several problems with short-rate models just because of their simplicity. First, the movements of yields and hence bond returns across maturities are perfectly correlated in short-rate models, which is contradicted by the empirical evidence. Second, although the short-rate models can accommodate monotonic increasing/decreasing and humped shapes of the yield curve, an inversely humped yield curve cannot be generated with these models. Moreover, with time-invariant parameters, short-rate models provide poor fit to the actual yield curves observed in market. This is a more severe problem in the pricing of fixed income derivatives as a close fit of the current yield curve is often required, which makes the short-rate models with time-invariant parameters less useful. To overcome this problem, one can constantly recalibrate the parameters in fitting the current yield curve. However, the repeated calibration of parameters clearly violates the assumption of parameter consistency which

could result the possibility of arbitrage opportunities. A better approach is to consider models with time-varying parameters instead of constant ones. Without further restrictions on model parameters, a perfect fit to the initial yield curve could always be obtained, yet the future yield curves are still entirely determined by the single state variable. Another alternative approach to solve the problems in simple short-rate models is to incorporate multiple factors in a term structure model, which is more flexible in generating richer yield curve dynamics.

Multi-factor affine models assume the evolution of the yield curve over time depends on several factors (e.g. Brennan and Schwartz (1979), Longstaff and Schwartz (1992)). Chen (1996) built a three-factor model with factors including the short rate, the mean and the volatility of the short rate. Duffie and Kan (1996) generalize a multi-factor affine term structure model, which nests many of the single and multi-factor models in the existing literature. In addition, they proposed to explain yields with latent factors that is not observable in market but can be inverted from yields. Dai and Singleton (2000) later tested the specification of these models and classified these models by counting the number m of processes that enter the volatility. In notation, $A_m(N)$ denotes a model with a total of N state variables, of which m enter the volatility.

Duffie and Kan's generalization of affine term structure models were originally derived in continuous time. To write their model in discrete time, we consider a general case of the yield curve driven by m random factors stacked in column vector $X_t \in \mathbb{R}^m$, with $t \in \{0, 1, \dots, T\}$. The m -factor discrete-time affine model is then described by the following equations

$$X_t = \mu + \Phi X_{t-1} + \Sigma \varepsilon_t \quad (\text{III.46})$$

with $\mu \in \mathbb{R}^m$, $\Phi \in \mathbb{R}^{m \times m}$, $\Sigma_t \in \mathbb{R}^{m \times m}$ and ε_t is the m -by-1 standardized error terms with zero mean and unit variance. The short rate r_t is assumed to be a linear function of the factors such that $r_t = \alpha + \beta' X_t$.

The assumption of no-arbitrage guarantees that there is a risk-neutral measure Q , under which the price of any security that does not pay out dividends at time $t+1$ (for example, the zero-coupon bond) should have the following discounting equation

$$P_t = E_t^Q \left(\exp(-r_t) P_{t+1} \right)$$

To convert the risk-neutral measure to the true data generating measure (i.e. the physical measure), the Radon-Nikodym derivative will do the job. Denoting ξ_t as the Radon-Nikodym derivative, for any random variable G_t , we have the following relation:

$$E_t^Q(G_{t+1}) = \frac{E_t(\xi_{t+1} G_{t+1})}{\xi_t}$$

If we assume that ξ_t follows the log-normal process as

$$\xi_{t+1} = \xi_t \exp\left(-\frac{1}{2}\Lambda_t' \Lambda_t - \Lambda_t' \varepsilon_{t+1}\right) \quad (\text{III.47})$$

where Λ_t is the time-varying market price of risk associated with the uncertainty in economy ε_t as $\Lambda_t = \lambda_0 + \lambda_1' X_t$. The pricing kernel m_t then could be defined as

$$M_{t+1} = \frac{\exp(-r_t) \xi_{t+1}}{\xi_t}$$

By substituting in the processes for short-rate and ξ_{t+1} , we now can link the pricing kernel with the m factors as

$$M_{t+1} = \exp\left(-\alpha - \beta' X_t - \frac{1}{2}\Lambda_t' \Lambda_t - \Lambda_t' \varepsilon_{t+1}\right)$$

If we let $P_t^{(n)}$ be the price of a zero coupon bond with maturity n , we can price the bond using the very intuitive no-arbitrage pricing equation $P = E(MX)$, where P is asset price, X is the pay-off of the asset and M is the pricing kernel. Now, guessing⁹ the price of the zero coupon bond is related to the m factors in an exponential affine fashion as

$$P_{n,t} = \exp(A_n + B_n' X_t) \quad (\text{III.48})$$

Then, from the recursive substitution, we derive the general expression of the zero coupon bond price as

⁹ This can be verified for the one-period bond case, in which we have $P_t^{(1)} = E_t(M_{t+1} \times 1) = \exp(-r_t) = \exp(\alpha + \beta' X_t)$

$$\begin{aligned}
P_t^{(n+1)} &= E_t \left(M_{t+1} P_{t+1}^{(n)} \right) \\
&= E_t \left(\exp \left(-\alpha - \beta' X_t - \frac{1}{2} \Lambda_t' \Lambda_t - \Lambda_t' \varepsilon_{t+1} + A_n + B_n' X_{t+1} \right) \right) \\
&= \exp \left(-\alpha - \beta' X_t - \frac{1}{2} \Lambda_t' \Lambda_t + A_n \right) E_t \left(\exp \left(-\Lambda_t' \varepsilon_{t+1} + B_n' X_{t+1} \right) \right) \\
&= \exp \left(-\alpha - \beta' X_t - \frac{1}{2} \Lambda_t' \Lambda_t + A_n \right) E_t \left(\exp \left(-\Lambda_t' \varepsilon_{t+1} + B_n' (\mu + \Phi X_t + \Sigma \varepsilon_{t+1}) \right) \right) \quad (III.49) \\
&= \exp \left(-\alpha + (B_n' \Phi - \beta') X_t - \frac{1}{2} \Lambda_t' \Lambda_t + A_n + B_n' \mu \right) E_t \left(\exp \left(-\Lambda_t' \varepsilon_{t+1} + B_n' \Sigma \varepsilon_{t+1} \right) \right) \\
&= \exp \left(-\alpha + (B_n' \Phi - \beta') X_t + A_n + B_n' (\mu - \Sigma \lambda_0) + \frac{1}{2} B_n' \Sigma \Sigma' B_n - B_n' \Sigma \lambda_1 X_t \right)
\end{aligned}$$

here in the last step, we use the log-normality to take out the expectation operator. By matching coefficients of (III.48) and (III.49), now we have the two expressions for A_{n+1} and B_{n+1}' as

$$\begin{aligned}
A_{n+1} &= -\alpha + A_n + B_n' (\mu - \Sigma \lambda_0) + \frac{1}{2} B_n' \Sigma \Sigma' B_n \\
B_{n+1}' &= B_n' (\Phi - \Sigma \lambda_1) - \beta'
\end{aligned}$$

III.5.3.1 Discussion on the specification of market price of risk in affine term structure models

The general solution of the affine term structure models given by Duffie and Kan (1996) is derived under risk-neutral measure. Thereafter, various extensions on the model are obtained with different specifications of the market price of risk. Simple specifications, like in Duffie and Kan (1996), are motivated by models' tractability in generating closed form solutions. By assigning each state variable a market price of risk that is proportional to the square root of that state variable, the product of the market price of risk and volatility is proportional to the state variables themselves. As a result, the drift term of the state variable dynamics is affine under both physical and risk-neutral measures. This justifies the popularity of the square-root process in many financial economic applications (e.g. term structure model of Cox, *et al.* (1985) and stochastic volatility model of Heston (1993)), Duffee (2002) called this class of affine models as the completely affine models. However, this specification of the market price of risk is rather restrictive, because this does not allow the sign of market price of risk to change which may result a downward bias on forecast performance. In his paper, Duffee (2002) generalizes the complete affine models to the essentially affine models by extending the market price of risk to include the inverse of the square-root of the state

variables while maintaining the affinity of the state variable dynamics under both measures. We will utilise this essentially affine term structure model in Chapter V to jointly model the nominal and real term structure of UK interest rates.

Bibliography

- Ang, Andrew, Geert Bekaert, and Min Wei, 2008a, The term structure of real rates and expected inflation, 63, 797-849.
- Balduzzi, Pierluigi, Giuseppe Bertola, and Silverio Foresi, 1997, A model of target changes and the term structure of interest rates, *Journal of Monetary Economics* 39, 223-249.
- Brennan, Michael J., and Eduardo S. Schwartz, 1979, A continuous time approach to the pricing of bonds, *Journal of Banking & Finance* 3, 133-155.
- Campbell, John Y., 1995, Some lessons from the yield curve, *The Journal of Economic Perspectives* 9, 129-152.
- Campbell, John Y., and Robert J. Shiller, 1991, Yield spreads and interest rate movements: A bird's eye view, *Review of Economic Studies* 58, 495.
- Chan, K. C., G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *The Journal of Finance* 47, 1209-1227.
- Chen, Lin, 1996, Stochastic mean and stochastic volatility - a three-factor model of the term structure of interest rates and its application to the pricing of interest rate derivatives, *Financial Markets, Institutions, and Instruments* 5, 1-88.
- Cochrane, John H., 2005. *Asset pricing* (Princeton University Press, Princeton, N.J.).
- Cochrane, John H., and Monika Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138-160.
- Cochrane, John H., and Monika Piazzesi, 2008, Decomposing the yield curve, *Working Paper*.
- Cox, John C., Jonathan E. Ingersoll, Jr., and Stephen A. Ross, 1985, An intertemporal general equilibrium model of asset prices, *Econometrica* 53, 363-384.
- Dai, Qiang, and Kenneth J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943-1978.
- Dai, Qiang, and Kenneth Singleton, 2003, Term structure dynamics in theory and reality, *Review of Financial Studies* 16, 631-678.

- Diebold, Francis X., and Canlin Li, 2006, Forecasting the term structure of government bond yields, *Journal of Econometrics* 130, 337-364.
- Duffee, Gregory R., 2002, Term premia and interest rate forecasts in affine models, *The Journal of Finance* 57, 405-443.
- Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379-406.
- Fama, Eugene F., and Robert R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680.
- Froot, Kenneth A., 1989, New hope for the expectations hypothesis of the term structure of interest rates, *The Journal of Finance* 44, 283-305.
- Hardouvelis, Gikas A., 1994, The term structure spread and future changes in long and short rates in the G7 countries: Is there a puzzle?, *Journal of Monetary Economics* 33, 255-283.
- Heston, Steven L. , 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* 6, 327-343.
- Kessler, Stephan and Scherer, Bernd (2009), Varying risk premia in international bond markets, *Journal of Banking and Finance*, 33-8, 1361-1375
- Lagarias, Jeffrey C. , James A. Reeds, Margaret H. Wright, and Paul E. Wright, 1998, Convergence properties of the nelder-mead simplex method in low dimensions, *SIAM Journal of Optimization* 9, 112-147.
- Litterman, Robert, and Jose A. Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income*.
- Longstaff, Francis, A., and Eduardo Schwartz, S., 1992, Interest rate volatility and the term structure: A two-factor general equilibrium model, *Journal of Finance* 47, 1259-1282.
- Mankiw, N. Gregory, and Jeffrey A. Miron, 1986, The changing behavior of the term structure of interest rates, *Quarterly Journal of Economics* 101, 211-228.
- Nelson, Charles R., and Andrew F. Siegel, 1987, Parsimonious modeling of yield curves, *The Journal of Business* 60, 473-489.
- Shiller, Robert J, and J Huston McCulloch, eds., 1990. *The term structure of interest rates*.
- Singleton, Kenneth, 2006. *Empirical dynamic asset pricing* (Princeton University Press).
- Thornton, Daniel L., 2006, Tests of the expectations hypothesis: Resolving the campbell-shiller paradox, *Journal of Money, Credit & Banking* 38, 511-542.

Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177-188.

Appendix III

Table III-2: Descriptive statistics

n	UK			US			German		
	$E(y_t^{(n)})$	$E(r_{t+1}^{(n)} - y_t^{(1)})$	Sharp Ratio	$E(y_t^{(n)})$	$E(r_{t+1}^{(n)} - y_t^{(1)})$	Sharp Ratio	$E(y_t^{(n)})$	$E(r_{t+1}^{(n)} - y_t^{(1)})$	Sharp Ratio
1	7.939(3.183)			6.073(2.860)			5.435(2.490)		
2	8.110(3.087)	0.483(1.872)	0.258	6.283(2.787)	0.487(1.860)	0.262	5.677(2.346)	0.666(1.651)	0.403
3	8.251(3.057)	0.850(3.384)	0.251	6.457(2.697)	0.843(3.404)	0.248	5.906(2.254)	1.283(3.009)	0.426
4	8.359(3.054)	1.095(4.790)	0.229	6.598(2.639)	1.095(4.713)	0.232	6.095(2.182)	1.741(4.147)	0.42
5	8.447(3.067)	1.303(6.080)	0.214	6.687(2.580)	1.123(5.771)	0.195	6.248(2.120)	2.092(5.148)	0.406
6	8.515(3.071)	1.490(7.255)	0.205				6.372(2.066)	2.364(6.058)	0.39
7	8.580(3.097)	1.650(8.355)	0.197				6.473(2.018)	2.584(6.912)	0.374
8	8.634(3.126)	1.771(9.426)	0.188				6.557(1.975)	2.767(7.730)	0.358
9	8.678(3.154)	1.869(10.513)	0.178				6.626(1.937)	2.930(8.525)	0.344
10	8.712(3.181)	1.942(11.654)	0.167				6.685(1.903)	3.079(9.316)	0.331

Note: $E(y_t^{(n)})$ is the average yield for a bond with a maturity of n; $E(r_{t+1}^{(n)} - y_t^{(1)})$ is the average holding period return in excess of one-period short rate (one-period bond yield) for a bond with maturity of n; sharp ratio is calculated as the ratio of average excess holding period return and the standard deviation of it; the standard deviations are shown in parentheses. In the empirical study, one period holding return is defined as 1 year holding return. The UK zero-coupon yields data contains yields with maturities 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 years starting from January 1970 to September 2009. The US zero-coupon bond price data (Fama-Bliss ZCB price data) has maturities vary from 1 year to 5 years with one year increment. The sample period for the US data begins from January 1964 and ends at September 2009. The German zero-coupon bond yield data has maturities vary from 1 year to 10 years with one year increment. The sample period for German data starts from September 1972 to September 2009.

Table III-3: Fama-Bliss regression and "complementary regression" results

Panel A				UK				US				German			
n	a^*	b^*	\bar{R}^2	χ_1^2	a^*	b^*	\bar{R}^2	χ_1^2	a^*	b^*	\bar{R}^2	χ_1^2			
2	-0.429(0.275)	0.826(0.253)	0.137	10.676[0.001]	-0.142(0.265)	0.167(0.235)	0.003	0.505[0.477]	-0.506(0.287)	0.663(0.288)	0.089	5.319[0.021]			
3	-0.797(0.391)	1.066(0.280)	0.254	14.509[0.000]	-0.494(0.556)	0.529(0.359)	0.048	2.170[0.141]	-1.208(0.475)	0.898(0.337)	0.153	7.088[0.008]			
4	-1.020(0.482)	1.102(0.308)	0.295	12.785[0.000]	-0.870(0.606)	0.789(0.269)	0.123	8.591[0.003]	-2.070(0.417)	1.271(0.309)	0.298	16.943[0.000]			
5	-1.482(0.518)	1.294(0.268)	0.426	23.304[0.000]	-0.979(0.616)	0.879(0.238)	0.157	13.668[0.000]	-2.939(0.369)	1.614(0.218)	0.504	54.902[0.000]			
6	-1.665(0.479)	1.113(0.180)	0.441	38.328[0.000]					-3.338(0.333)	1.672(0.197)	0.649	72.208[0.000]			
7	-1.798(0.491)	0.972(0.153)	0.400	40.393[0.000]					-3.151(0.328)	1.472(0.211)	0.638	48.500[0.000]			
8	-2.169(0.474)	1.042(0.178)	0.474	34.308[0.000]					-2.778(0.444)	1.169(0.209)	0.547	31.338[0.000]			
9	-2.260(0.487)	0.959(0.232)	0.426	17.031[0.000]					-2.364(0.528)	0.787(0.182)	0.36	18.697[0.000]			
10	-2.370(0.504)	0.878(0.223)	0.396	15.504[0.000]					-2.349(0.505)	0.606(0.155)	0.308	15.299[0.000]			
Panel B				UK				US				German			
n	a^c	b^c	\bar{R}^2	χ_1^2	a^c	b^c	\bar{R}^2	χ_1^2	a^c	b^c	\bar{R}^2	χ_1^2			
2	0.429(0.275)	0.174(0.253)	0.005	0.471[0.492]	0.142(0.265)	0.833(0.235)	0.112	12.609[0.000]	0.506(0.287)	0.337(0.288)	0.023	1.368[0.242]			
3	0.790(0.471)	0.110(0.348)	-0.001	0.101[0.751]	0.037(0.492)	1.131(0.313)	0.126	13.093[0.000]	0.824(0.589)	0.502(0.385)	0.031	1.706[0.192]			
4	1.087(0.640)	0.012(0.436)	-0.002	0.001[0.979]	-0.165(0.674)	1.366(0.388)	0.140	12.375[0.000]	0.947(0.866)	0.662(0.450)	0.041	2.161[0.142]			
5	1.360(0.801)	-0.072(0.490)	-0.002	0.022[0.883]	0.164(0.894)	1.046(0.470)	0.062	4.946[0.026]	0.996(1.115)	0.787(0.512)	0.047	2.362[0.124]			
6	1.549(0.942)	-0.066(0.517)	-0.002	0.016[0.898]					1.031(1.328)	0.878(0.573)	0.048	2.352[0.125]			
7	1.633(1.073)	0.017(0.533)	-0.002	0.001[0.974]					1.029(1.515)	0.971(0.626)	0.05	2.405[0.121]			
8	1.624(1.206)	0.146(0.543)	-0.001	0.073[0.788]					1.007(1.673)	1.064(0.674)	0.052	2.488[0.115]			
9	1.564(1.353)	0.300(0.550)	0.002	0.297[0.586]					0.991(1.800)	1.146(0.717)	0.053	2.550[0.110]			
10	1.468(1.518)	0.473(0.548)	0.007	0.743[0.389]					0.970(1.898)	1.227(0.750)	0.054	2.678[0.102]			

Note: Panel A shows the results from regression $y_{t+n-1}^{(1)} - y_t^{(1)} = a^* + b^* (f_t^{(n)} - y_t^{(1)}) + \varepsilon_t$. Panel B shows the results from regression $r_{t+1}^{(n)} - y_t^{(1)} = a^c + b^c (f_t^{(n)} - y_t^{(1)})$. The standard errors, which are shown in parentheses, are calculated using Newey-West method with 24 lags to correct the data overlapping problem. The p-values for χ_1^2 statistics are shown in square brackets.

Table III-4: Time series and cross sectional regressions of 1-year excess bond returns on all forward rates

Panel A													
	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}	\bar{R}^2	$\chi^2_{5/10}$
UK	-3.068 (2.256)	-0.699 (1.263)	3.785 (7.201)	-5.342 (15.799)	4.253 (9.405)	-6.520 (14.432)	9.297 (10.179)	5.698 (10.418)	-10.456 (6.694)	-8.258 (6.574)	8.820 (4.483)	0.046	7.434 [0.684]
US	-2.088 (1.300)	-1.591 (0.449)	0.045 (0.665)	2.434 (0.732)	1.213 (0.521)	-1.815 (0.438)						0.261	45.207 [0.000]
German	-3.012 (2.888)	-1.474 (0.856)	2.783 (1.819)	-8.147 (6.213)	3.773 (7.138)	19.936 (9.538)	-21.871 (9.735)	-3.412 (6.409)	11.967 (7.556)	3.650 (4.650)	-6.639 (4.288)	0.149	50.876 [0.000]
Panel B													
	Restricted						Unrestricted						
	UK		US		German		UK		US		German		
n	$\beta^{(n)}$	\bar{R}^2	$\beta^{(n)}$	\bar{R}^2	$\beta^{(n)}$	\bar{R}^2	\bar{R}^2	χ^2_{10}	\bar{R}^2	χ^2_5	\bar{R}^2	χ^2_{10}	
2	0.303 (0.129)	0.066	0.465 (0.075)	0.236	0.263 (0.061)	0.098	0.078	10.887 [0.366]	0.240	48.218 [0.000]	0.107	50.253 [0.000]	
3	0.563 (0.219)	0.078	0.867 (0.144)	0.255	0.517 (0.104)	0.121	0.087	14.560 [0.149]	0.252	46.046 [0.000]	0.122	55.066 [0.000]	
4	0.779 (0.304)	0.082	1.244 (0.208)	0.286	0.729 (0.134)	0.140	0.086	15.781 [0.106]	0.280	54.293 [0.000]	0.132	52.715 [0.000]	
5	0.961 (0.380)	0.081	1.424 (0.259)	0.259	0.910 (0.159)	0.155	0.076	13.831 [0.181]	0.255	38.136 [0.000]	0.141	49.616 [0.000]	
6	1.109 (0.446)	0.077			1.066 (0.181)	0.165	0.062	10.635 [0.387]			0.147	47.155 [0.000]	
7	1.223 (0.503)	0.070			1.203 (0.203)	0.171	0.051	8.137 [0.615]			0.152	47.317 [0.000]	
8	1.302 (0.559)	0.062			1.328 (0.224)	0.174	0.043	6.870 [0.738]			0.156	49.116 [0.000]	
9	1.357 (0.618)	0.053			1.441 (0.245)	0.175	0.040	7.331 [0.694]			0.159	50.752 [0.000]	
10	1.403 (0.683)	0.046			1.543 (0.266)	0.172	0.043	9.455 [0.490]			0.161	53.782 [0.000]	

Note: Standard errors, which are shown in parentheses, are calculated as Newey-West standard errors with 24 lags. The p-values for χ^2_5 and χ^2_{10} statistics are shown in square brackets.

Table III-5: Time series and cross sectional regressions of 2-year excess bond returns on all forward rates

Panel A	$\tilde{\gamma}_0^{SF}$	$\tilde{\gamma}_1^{SF}$	$\tilde{\gamma}_2^{SF}$	$\tilde{\gamma}_3^{SF}$	$\tilde{\gamma}_4^{SF}$	$\tilde{\gamma}_5^{SF}$	$\tilde{\gamma}_6^{SF}$	$\tilde{\gamma}_7^{SF}$	$\tilde{\gamma}_8^{SF}$	$\tilde{\gamma}_9^{SF}$	$\tilde{\gamma}_{10}^{SF}$	\bar{R}^2	$\chi_{5/10}^2$
UK	-2.248 (0.489)	0.034 (0.225)	-0.558 (0.752)	3.242 (2.038)	-6.184 (2.814)	2.291 (1.884)	3.510 (2.088)	0.715 (1.441)	-2.954 (1.505)	-1.861 (0.994)	2.159 (0.914)	0.456	166.686 [0.000]
US	-3.166 (0.456)	-0.070 (0.203)	-1.127 (0.496)	1.096 (0.434)	0.189 (0.196)	0.420 (0.202)						0.374	116.03 [0.000]
German	-4.523 (1.152)	0.107 (0.194)	0.571 (0.884)	-4.138 (2.190)	3.920 (2.815)	3.007 (2.720)	-4.966 (2.340)	-1.128 (2.085)	2.761 (2.751)	1.657 (1.451)	-0.921 (1.259)	0.422	74.014 [0.000]
Panel B	Restricted						Unrestricted						
	UK		US		German		UK		US		German		
n	$\tilde{\beta}^{SF,(n)}$	\bar{R}^2	$\tilde{\beta}^{SF,(n)}$	\bar{R}^2	$\tilde{\beta}^{SF,(n)}$	\bar{R}^2	\bar{R}^2	χ_{10}^2	\bar{R}^2	χ_5^2	\bar{R}^2	χ_{10}^2	
2	0.216 (0.106)	-0.011	0.428 (0.126)	0.052	0.305 (0.100)	0.054	0.064	14.247 [0.162]	0.101	10.111 [0.072]	0.174	41.310 [0.000]	
3	0.361 (0.180)	-0.015	0.679 (0.229)	0.030	0.573 (0.181)	0.052	0.056	14.323 [0.159]	0.064	6.132 [0.294]	0.152	36.772 [0.000]	
4	0.484 (0.254)	-0.011	0.934 (0.316)	0.036	0.781 (0.244)	0.051	0.049	13.588 [0.193]	0.057	5.932 [0.313]	0.140	32.356 [0.000]	
5	0.620 (0.329)	-0.007	1.027 (0.383)	0.035	0.953 (0.299)	0.054	0.047	12.452 [0.256]	0.053	5.462 [0.362]	0.137	30.518 [0.001]	
6	0.766 (0.402)	-0.002			1.104 (0.347)	0.059	0.046	11.584 [0.314]			0.138	30.446 [0.001]	
7	0.911 (0.470)	0.002			1.238 (0.393)	0.063	0.046	11.290 [0.335]			0.137	30.471 [0.001]	
8	1.045 (0.536)	0.006			1.363 (0.435)	0.067	0.047	11.702 [0.305]			0.135	30.737 [0.001]	
9	1.171 (0.599)	0.010			1.480 (0.475)	0.070	0.047	12.811 [0.234]			0.131	30.388 [0.001]	
10	1.292 (0.662)	0.012			1.591 (0.513)	0.073	0.048	14.258 [0.162]			0.125	29.951 [0.001]	

Note: Standard errors, which are shown in parentheses, are calculated as Newey-West standard errors with 24 lags. The p-values for χ_5^2 and χ_{10}^2 statistics are shown in square brackets

Matlab codes used for Nelson-Siegel factor model (Kalman filter approach)

The following codes require CompEcon Toolbox to run. The toolbox can be downloaded from www4.ncsu.edu/~pfackler/compecon

```
%% =====NS_kalman.m=====
%% This is the main code file, it calls the NS_kalman_likfun.m to minimize
%% the negative value of the likelihood value, calls the NS_filter.m to
%% filter out the unobserved NS factors
%% =====
% Nelson-Siegel kalman filter estimation (one-step)
% model:
% ----state equation----
% NS(t)=alpha+beta*NS(t-1)+Q*e(t)
%
% ----measurement equation----
% Y(t)=loadings'*NS(t)+u(t)
% => Y(t)=TT*NS(t)+u(t)
% where NS(t)=[NS_level;
%              NS_slope;
%              NS_curve;]
% and loadings=[1;
%               (1-exp(-lambda*n))/(lambda*n);
%               (1-exp(-lambda*n))/(lambda*n)-exp(-lambda*n)];
%
% data: German ZCB monthly yields 1972:09~2009:09 1-year to 10-year

clear all;
close all;

% global variables
global Y yields_names yields_maturity yields_obs yields_length
global num_factors Q loadings
%% data preparation and parameter defination
% datapath='GM_Monthly_0972to0909.mat';
% load(datapath);
% yields=GM_Monthly_0972to0909(:,4:end)./1200;
% date_t=9/12+1972:1/12:9/12+2009;
data=load('UK1to10Y_Jan70_Sep09.txt');
yields=data./1200;
date_t=1/12+1970:1/12:9/12+2009;
% parameters
yields_names=['1y' '2y' '3y' '4y' '5y' '6y' '7y' '8y' '9y' '10y'];
yields_maturity=[12 24 36 48 60 72 84 96 108 120]; % in month
yields_obs=size(yields,2);
yields_length=size(yields,1);
lambda=0.077;
num_factors=3;
lag=1;
rolling=1;

%% NS loadings
loading_L =ones(1,yields_obs);
loading_SL =zeros(1,yields_obs);
loading_C =zeros(1,yields_obs);
for i1=1:yields_obs
    loading_SL(i1)=(1-exp(-lambda*yields_maturity(i1)))/(lambda*yields_maturity(i1));
    loading_C(i1) =loading_SL(i1)-exp(-lambda*yields_maturity(i1));
end
```

```

loadings=[loading_L;
           loading_SL;
           loading_C]; % 3 by 10
%%% construct system matrices for measurement equation and state equation
Y=yields;
TT=loadings';
Q=eye(num_factors);
%%% invert NS-factors
NS=((loadings*loadings')\eye(size(loadings,1)))*loadings*Y'; % 3 by 403

%%% initial values guess for state equation (VAR(1) estimation)
yy=NS(:,lag+1:end);
xx=[ones(1,yields_length-1);
    NS(:,lag:end-1)];
phi_in=(yy*xx')/(xx*xx');
alpha_in=phi_in(:,1);
beta_in=phi_in(:,2:4);
% variance for state
e_in=(yy-phi_in*xx)*(yy-phi_in*xx)/(yields_length-lag);
%var for measurement equation
u_in=((Y(1+lag:end,:))'-TT*NS(:,1+lag:end))*((Y(1+lag:end,:))'-TT*NS(:,1+lag:end))/(yields_length-lag);
e_in2=[e_in(1,1); e_in(1,2); e_in(1,3);e_in(2,2); e_in(2,3); e_in(3,3)];
Theta=[alpha_in;
        reshape(beta_in,num_factors^2,1);
        e_in2;
        diag(u_in)];
%%%
options=optimset('Display','iter','TolX',1e-4,'TolFun',1e-8,...
                 'MaxFunEvals',1000000,'MaxIter',1000000);
[Theta_hat,fval,exitflag,output]=fminsearch(@NS_kalman_likfun,Theta,options);
%%%
xfnl=para_trans(Theta_hat);
% S.E. and t-test
h_0=fdhess(@NS_kalman_likfun,Theta_hat); % calculate Hessian matrix
g_0=fdjac1(@para_trans,Theta_hat,[]); % calculate gradient
h_fnl=g_0*(inv(h_0))*g_0'; % final Hessian
std_fnl=sqrt(diag(h_fnl)); % S.E.
t_ratio=xfnl/std_fnl; % t-ratio

%%% plot the NS factors, factor residuals, fitted yields and residuals of yields
[store_NS,store_u,store_fit]=NS_filter(Theta_hat);

function minus_loglikvalue=NS_kalman_likfun(para_in)
% This function evaluates the NS model likelihood value function
global Y yields_names yields_maturity yields_obs yields_length
global num_factors Q loadings

para=para_trans(para_in);

alpha=para(1:3);
beta=reshape(para(4:12),3,3);
e=[para(13) para(14) para(15);
   para(14) para(16) para(17);
   para(15) para(17) para(18)];
sq_u=[para(19) 0 0 0 0 0 0 0 0 0;
      0 para(20) 0 0 0 0 0 0 0 0;
      0 0 para(21) 0 0 0 0 0 0 0;
      0 0 0 para(22) 0 0 0 0 0 0;

```



```

0      0      0      0      para(23) 0      0      0      0      0;
0      0      0      0      0      para(24) 0      0      0      0;
0      0      0      0      0      0      para(25) 0      0      0;
0      0      0      0      0      0      0      para(26) 0      0;
0      0      0      0      0      0      0      0      para(27) 0;
0      0      0      0      0      0      0      0      0      para(28)];
% initial value of the likelihood function given the guessed parameters
likvalue=-0.5*yields_obs*(yields_length-1)*log(2*pi);
%likvalue=0;
% initial value of the NS factors
NS_0=(inv(eye(num_factors,num_factors)-beta))*alpha;
%%
NS_lag=NS_0;
vec_e=(eye(num_factors^2)-kron(beta,beta'))\eye(num_factors^2)*(reshape(e,9,1));
%vec_e=e;
e_lag=reshape(vec_e,num_factors,num_factors);
for t=1:yields_length
    NS_t = alpha+beta*NS_lag;
    e_t = beta*e_lag*beta'+e;
    err_t = Y(t,:)-loadings'*NS_t;
%    var_err_t = loadings'*e_t*loadings+sq_u*eye(yields_obs);
    var_err_t = loadings'*e_t*loadings+sq_u;
% likelihood function
    if t>1
        likvalue=likvalue-0.5*log(det(var_err_t))-0.5*err_t'*inv(var_err_t)*err_t;
    end
% updating
    K_gain = e_t*loadings*inv(var_err_t);
    NS_lag = NS_t+K_gain*err_t;
    e_lag = e_t-K_gain*loadings'*e_t;
%store_NS(:,t) = NS_lag;
%figure(1)
%plot(store_NS(1,:));
end
minus_loglikvalue=-likvalue;
if e<0;
    minus_loglikvalue = real(minus_loglikvalue) + 1e8;
end

if abs(imag(minus_loglikvalue)) > 0
    minus_loglikvalue = real(minus_loglikvalue) + 1e8;
end

```

Chapter IV A Markov switching extension of the short rate model

IV.1 Introduction

In recent years, Markov switching models have been prolifically applied to many financial time series data. Hamilton (1989b), Cai (1994), Gray (1996), Naik and Lee (1997) and Bansal and Zhou (2002) have conducted successful empirical studies on interest rate modelling using the Markov switching framework. These studies all report the existence of regime changes in the evolution of short rates. In particular, there is strong evidence supporting high and low volatilities in short rate dynamics over time.

In Cai (1994), the fitted model is a Markov switching ARCH model, which deliberately avoided the problem of infinite path dependency of using a Markov switching GARCH model (i.e. avoids the conditional variance at time t depending on the whole sample path). Gray (1996) overcame the path dependency problem by using the conditional expectation of past variances (i.e. integrating out the unobserved regime path in the GARCH equation), which has found both the level effect and GARCH specification of the errors are needed to model the dynamics of the U.S. short rates. Naik and Lee (1997) employed a variation of CIR model, but only allows the volatility of short rate to switch regimes. However, none of them compares the impacts of different model specifications on the models' pricing performance. One exception is Driffill, Kenc, Sola and Spagnolo (2009), where the authors evaluated different parameterizations of the Markov switching CIR model in terms of real-time one-step-ahead bond pricing performance.

In this study, we apply Driffill, *et al.* (2009)'s methodology in modelling the term structure of interest rates on the UK's government bonds data. We conduct a series of model specification tests on a battery of models that employ different parameter restrictions on the short rate equation. By doing this, we provide a comprehensive specification analysis of the Markov switching CIR model on the UK's data that varies from January 1970 to September 2010. In comparison with Driffill, *et al.* (2009), our model specification analysis covers each possible combinations of different regime switching parameters in the short rate equation. In addition, the sample period we studied in the UK market is far longer than their sample period in the US market, which ranges from January 1964 to April 1998.

We find that the least restricted model gives the best goodness-of-fit on the UK's term structure of interest rates. Although other simplified models may provide slightly better out-of-sample forecast in directional movements of the yields, the economic gains of employing simplified Markov switching Cox-Ingersoll-Ross (hereafter CIR) models on the UK term structure of interest rates is relatively small.

This study is organized as follows. In section 2, we review the basic single regime CIR model of short rates. In section 3, we discuss the Markov switching extension of the CIR model. Section 4 presents the empirical results of the Markov switching CIR model applied on the UK bond yields data. Section 5 evaluates the performance of the Markov switching CIR model by focusing on the ex ante pricing performance of various models. Finally, section 6 concludes.

IV.2 Single regime CIR model

In a single regime CIR model, the short rate r_t is driven by a single factor x_t , which follows a mean-reverting square root process. With a discrete-time setting, this single factor is written as

$$x_{t+1} - x_t = k(\delta - x_t) + \sigma\sqrt{x_t}\varepsilon_{t+1}, \quad (\text{IV.1})$$

where k is the speed of adjustment for x_{t+1} to revert back to its long-term equilibrium value δ , ε_{t+1} is a series of normally distributed shocks with zero mean and unit variance, and $\sigma^2 x_t$ is the unexpected shock to x_{t+1} , in which σ is the local volatility served as a scaling parameter and x_t is the level of the factor at time t . Since the majority of empirical evidence suggests that the conditional variance of short rate is time-varying, the CIR specification of short rate that allows its conditional variance to be linked with its level is more appealing when compared to the constant conditional variance specification in Vasicek (1977).

The pricing kernel, M_t , in a discrete-time CIR model is defined as

$$M_{t+1} = \exp\left(-r_t - \frac{x_t}{2}\left(\frac{\lambda}{\sigma}\right)^2 - \left(\frac{\lambda}{\sigma}\right)\sqrt{x_t}\varepsilon_{t+1}\right), \quad (\text{IV.2})$$

where λ is the market price of risk that determines the covariance of M_t and x_t . Market price of risk plays an important role in interest rates term structure modelling. In finance theory, the expected bond returns differ from the risk-free rate by a risk premium in a risk-averse world. This risk premium depends on the exposure to various sources of risks which is denoted as ε_t in the above pricing kernel. The market price of risk represents the compensation in the expected return for exposing one unit of risk introduced by ε_t . Each component of ε_t has its own market price of risk represented by the corresponding λ . Since CIR model considers one factor (i.e. the short rate) driving the whole term structure, the only source of risk comes from the ε_t in the short rate equation, hence the only market price of risk λ in the pricing kernel.

To price a zero coupon bond at time t with maturity n , we use the no-arbitrage condition implied fundamental pricing equation $P_t^{(n+1)} = E_t(M_{t+1}P_{t+1}^{(n)})$. By guessing that the bond price follows an affine functional form of $P_t^{(n+1)} = \exp(A_{n+1} + B_{n+1}x_t)$ and using the boundary condition $P_{t+n}^{(0)} = 1$, we have

$$\begin{aligned}
P_t^{(n+1)} &= E_t(M_{t+1}P_{t+1}^{(n)}) \\
&= E_t \left[\exp \left(-x_t - \frac{x_t}{2} \left(\frac{\lambda}{\sigma} \right)^2 - \left(\frac{\lambda}{\sigma} \right) \sqrt{x_t} \varepsilon_{t+1} + A_n + B_n (x_t + k(\delta - x_t) + \sigma \sqrt{x_t} \varepsilon_{t+1}) \right) \right] \\
&= \exp \left(-x_t - \frac{x_t}{2} \left(\frac{\lambda}{\sigma} \right)^2 + A_n + B_n x_t + B_n k(\delta - x_t) \right) E_t \left(\exp \left(- \left(\frac{\lambda}{\sigma} \right) \sqrt{x_t} \varepsilon_{t+1} + B_n \sigma \sqrt{x_t} \varepsilon_{t+1} \right) \right) \\
&= \exp \left(-x_t - \frac{x_t}{2} \left(\frac{\lambda}{\sigma} \right)^2 + A_n + B_n x_t + B_n k(\delta - x_t) \right) \exp \left(\frac{1}{2} \left(B_n \sigma - \left(\frac{\lambda}{\sigma} \right) \right)^2 x_t \right) \\
&= \exp \left(-x_t - \frac{x_t}{2} \left(\frac{\lambda}{\sigma} \right)^2 + A_n + B_n x_t + B_n k(\delta - x_t) \right) \exp \left(\left(\frac{1}{2} B_n^2 \sigma^2 + \frac{1}{2} \left(\frac{\lambda}{\sigma} \right)^2 - B_n \lambda \right) x_t \right) \\
&= \exp \left((A_n + B_n k \delta) + \left(-1 - \frac{1}{2} \left(\frac{\lambda}{\sigma} \right)^2 + B_n - B_n k + \frac{1}{2} B_n^2 \sigma^2 + \frac{1}{2} \left(\frac{\lambda}{\sigma} \right)^2 - B_n \lambda \right) x_t \right) \\
&= \exp \left((A_n + B_n k \delta) + \left(-1 + (1 - k - \lambda) B_n + \frac{1}{2} B_n^2 \sigma^2 \right) x_t \right) \\
&= \exp(A_{n+1} + B_{n+1} x_t).
\end{aligned}$$

Therefore,

$$A_{n+1} = A_n + B_n k \delta$$

$$B_{n+1} = -1 + (1 - k - \lambda) B_n + \frac{1}{2} B_n^2 \sigma^2.$$

Also note that the boundary condition $P_{t+n}^{(0)} = 1$ implies $A_0 = B_0 = 0$.

IV.3 Markov switching CIR model

In a Markov switching CIR model, the evolution of the factor that drives the short rate depends on an unobserved first-order Markov chain process S_t . Assume S_t has two states which take value of either 0 for regime 0 or 1 for regime 1. The regime switches are accounted by allowing parameters k , δ , and σ to be regime dependent, i.e. $k(S_t)$, $\delta(S_t)$ and $\sigma(S_t)$. The switches between regimes are governed by a transition probability matrix $p = (p_{ij})$ for $i, j = \{0, 1\}$, where $p_{ij} = \Pr(S_{t+1} = j | S_t = i)$ with $\sum_{j=0,1} p_{ij} = 1$ and $0 < p_{ij} < 1$. Additionally, it is assumed that S_t is independent of the random shocks ε_t and the econometrician does not know the actual state prevailing in the market but the investors know.

The factor process is now written as

$$x_{t+1} - x_t = k(j)(\delta(j) - x_t) + \sigma(j)\sqrt{x_t}\varepsilon_{t+1}. \quad (\text{IV.3})$$

Following Bansal and Zhou (2002), we allow the market price of risk to be regime dependent and accordingly adjust the pricing kernel as

$$M_{t+1}(j) = \exp \left[-r_t - \frac{x_t}{2} \left(\frac{\lambda(j)}{\sigma(j)} \right)^2 - \frac{\lambda(j)}{\sigma(j)} \sqrt{x_t} \varepsilon_{t+1} \right].$$

By guessing the zero coupon bond price at time t in regime i is $P_t^{(n+1)}(i) = \exp(A_{n+1}(i) + B_{n+1}(i)x_t)$, with initial conditions $A_0(i) = B_0(i) = 0$, the bond pricing equation, when subjecting to regime switches, now becomes

$$\begin{aligned}
P_t^{(n+1)}(i) &= \sum_{j=0,1} p_{ij} E_t \left(M_{t+1}(j) P_{t+1}^{(n)}(j) \right) \\
&= \sum_{j=0,1} p_{ij} E_t \left(\exp \left(-r_t - \frac{x_t}{2} \left(\frac{\lambda(j)}{\sigma(j)} \right)^2 - \frac{\lambda(j)}{\sigma(j)} \sqrt{x_t} \varepsilon_{t+1} \right) \exp(A_n(j) + B_n(j) x_{t+1}) \right) \\
&= \sum_{j=0,1} p_{ij} E_t \left(\exp \left(-x_t - \frac{x_t}{2} \left(\frac{\lambda(j)}{\sigma(j)} \right)^2 - \frac{\lambda(j)}{\sigma(j)} \sqrt{x_t} \varepsilon_{t+1} + A_n(j) + \right. \right. \\
&\quad \left. \left. B_n(j) [x_t + k(j)(\delta(j) - x_t) + \sigma(j) \sqrt{x_t} \varepsilon_{t+1}] \right) \right) \\
&= \sum_{j=0,1} p_{ij} \left[\exp \left(-x_t - \frac{x_t}{2} \left(\frac{\lambda(j)}{\sigma(j)} \right)^2 + A_n(j) + B_n(j) x_t + \right. \right. \\
&\quad \left. \left. B_n(j) k(j) \delta(j) - B_n(j) k(j) x_t + \frac{1}{2} \left(B_n(j) \sigma(j) - \frac{\lambda(j)}{\sigma(j)} \right)^2 x_t \right) \right] \\
&= \sum_{j=0,1} p_{ij} \left[\exp \left((A_n(j) + B_n(j) k(j) \delta(j)) + \right. \right. \\
&\quad \left. \left. \left(B_n(j) - B_n(j) k(j) + \frac{1}{2} B_n(j)^2 \sigma(j)^2 - B_n(j) \lambda(j) - 1 \right) x_t \right) \right] \\
&= \sum_{j=0,1} p_{ij} \left[\exp \left((A_n(j) + B_n(j) k(j) \delta(j)) + \right. \right. \\
&\quad \left. \left. \left((1 - k(j) - \lambda(j)) B_n(j) + \frac{1}{2} B_n(j)^2 \sigma(j)^2 - 1 \right) x_t \right) \right].
\end{aligned}$$

Thus, we have

$$\begin{aligned}
A_{n+1}(i) &= \sum_{j=0,1} p_{ij} (A_n(j) + B_n(j) k(j) \delta(j)) \\
B_{n+1}(i) &= \sum_{j=0,1} p_{ij} \left((1 - k(j) - \lambda(j)) B_n(j) + \frac{1}{2} B_n(j)^2 \sigma(j)^2 - 1 \right)
\end{aligned}$$

IV.3.1 Maximum Likelihood Estimation of Markov switching CIR model

Assume the yields are observed with errors, which yields

$$y_t^{(n)}(i) = -\frac{A_n(i)}{n} - \frac{B_n(i)}{n} x_t + u_t(i), \quad (\text{IV.4})$$

where $u_t(i)$ is normally distributed with zero mean and regime dependent variance-covariance matrix $\Sigma(i)$. Following Hamilton (1989b), the density of the yields, y_t , conditional on its history and state variable S_t and the parameter space $\Psi = \{k_0, k_1, \delta_0, \delta_1, \sigma_0, \sigma_1, \lambda_0, \lambda_1, \Sigma_0, \Sigma_1\}$, can be written as

$$\begin{aligned}
& P(y_t | y_{t-1}, \dots, y_1, S_t, \Psi) \\
&= \frac{1}{\sqrt{2\pi}\sigma(i)\sqrt{x_{t-1}}} \exp\left(-\frac{(x_t - x_{t-1} - k(i)(\delta(i) - x_{t-1}))^2}{2\sigma(i)^2 x_{t-1}}\right) \\
&\times \frac{1}{2\pi\sqrt{|\Sigma(i)|}} \exp\left(-\frac{u_t' u_t}{2\Sigma(i)}\right).
\end{aligned}$$

To estimate the model, we need to choose suitable data for the short rate that drives the whole term structure of UK yield curve. Since the theoretical instantaneous short rate does not exist in practice, it is often approximated by short-maturity yields. Like in Driffill, *et al.* (2009), we assume the UK's 3-month short rate as the instantaneous short rate. This ad hoc assumption is related to the widely recognised fact of poor coverage of the UK market at the short-end of the maturity spectrum. Evans (2003), for example, noted that the lack of nominal bonds with short maturities make it impossible to estimate the short end of the UK nominal yield curves precisely, and the estimated yields for one and two month bonds would contain significant sampling errors. In literatures studying the US term structure of interest rates, it is found that yields with maturities less than 3 month behave differently from other short rates and do not share much variation with other short-term yields, e.g. see Piazzesi (2010) and Duffee (1996).

The other two bond yields included in the estimation of (IV.4) are the UK's 5-year and 10-year bond yields. We assume the 3-month short rate drives 5-year and 10-year bond yields. This is because we consider not only the best fittings for the short-end of the yield curve but also the best fittings for mid-term and long-end of the yield curve. However, the impact on the fittings of the yield curve of choosing yields other than 5-year and 10-year is unclear, and remains an interesting one for future research.

All three yields included in estimation are assumed to be measured imperfectly. In other words, the 5-year and 10-year bond yields are priced as in equation (IV.4). We follow Driffill, *et al.* (2009) to assess the impacts on bond pricing from assuming different constraints on the parameters that govern the dynamics of the instantaneous short rate (in our case, the 3-month short rate). More specifically, we estimate a battery of models as shown in Table IV-1, where the differences between models are various assumptions on the regime dependence of the parameters in equation (IV.3). In comparison with Driffill, *et al.* (2009), which the authors only estimated Model 0,1,2,3 and 5 as shown in Table IV-1, our model specification analysis is more comprehensive. We provide analysis on a full range of models, under a Markov

switching CIR framework, covering each possible combinations of different regime switching parameters in equation (IV.3). In addition, the sample period we studied in the UK market is much longer than their sample period for the US market (which ranges from January 1964 to April 1998).

Table IV-1: Estimated models

Model 0: Single regime short rate specification

$$x_{t+1} - x_t = k(\delta - x_t) + \sigma\sqrt{x_t}\varepsilon_{t+1}$$

Model 1: Regime switching in all parameters of the short rate equation

$$x_{t+1} - x_t = k(j)(\delta(j) - x_t) + \sigma(j)\sqrt{x_t}\varepsilon_{t+1}$$

Model 2: Regime switching in all parameters of the short rate equation, except k

$$x_{t+1} - x_t = k(\delta(j) - x_t) + \sigma(j)\sqrt{x_t}\varepsilon_{t+1}$$

Model 3: Regime switching in all parameters of the short rate equation, except δ

$$x_{t+1} - x_t = k(j)(\delta - x_t) + \sigma(j)\sqrt{x_t}\varepsilon_{t+1}$$

Model 4: Regime switching in all parameters of the short rate equation, except σ

$$x_{t+1} - x_t = k(j)(\delta(j) - x_t) + \sigma\sqrt{x_t}\varepsilon_{t+1}$$

Model 5: Regime switching in all parameters of the short rate equation, except k and δ

$$x_{t+1} - x_t = k(\delta - x_t) + \sigma(j)\sqrt{x_t}\varepsilon_{t+1}$$

Model 6: Regime switching in all parameters of the short rate equation, except k and σ

$$x_{t+1} - x_t = k(\delta(j) - x_t) + \sigma\sqrt{x_t}\varepsilon_{t+1}$$

Model 7: Regime switching in all parameters of the short rate equation, except δ and σ

$$x_{t+1} - x_t = k(j)(\delta - x_t) + \sigma\sqrt{x_t}\varepsilon_{t+1}$$

IV.4 Empirical results

Table IV-2 displays the descriptive statistics of the UK's nominal yields with different maturities. As we can see from the table, one important feature of the data set is the existence of positive and high correlations between nominal yields with different maturities. This is consistent with many previous empirical studies on the yield curves, and suggests that a small

number of common factors drive the co-movement of the nominal yields across different maturities. However, as the correlation coefficients are not exactly unity, it means that there are non-parallel shifts which cannot be simply captured by a linear one-factor model. We mitigate this problem by incorporating a second but unobserved factor, i.e. the Markov switching variable, into the simple one-factor CIR model, which will enable us to generate non-normal distributions and nonlinearities in the conditional first and second moments.

Table IV-2: Descriptive statistics

	$y^{(3m)}$	$y^{(1y)}$	$y^{(2y)}$	$y^{(3y)}$	$y^{(4y)}$	$y^{(5y)}$	$y^{(6y)}$	$y^{(7y)}$	$y^{(8y)}$	$y^{(9y)}$	$y^{(10y)}$
Mean	6.717	6.568	6.618	6.689	6.750	6.798	6.837	6.865	6.883	6.891	6.889
Median	5.763	5.850	6.180	6.238	6.302	6.254	6.171	6.133	6.229	6.200	6.249
Maximum	14.67	14.31	13.72	13.31	13.07	12.93	12.82	12.73	12.62	12.50	12.36
	2	1	5	5	5	2	8	3	9	8	8
Minimum	0.403	0.579	0.771	1.080	1.422	1.757	2.070	2.353	2.606	2.828	3.022
Std. Dev.	3.463	3.138	2.887	2.740	2.652	2.597	2.558	2.528	2.501	2.473	2.443
Skewness	0.532	0.332	0.238	0.225	0.241	0.264	0.285	0.300	0.309	0.312	0.310
Kurtosis	2.763	2.654	2.409	2.205	2.054	1.941	1.852	1.779	1.719	1.667	1.621
Normality Test											
Jarque-Bera	15.26	7.220	7.421	10.76	14.52	18.04	21.15	23.81	26.05	27.90	29.44
	9			1	4	8	6	8	3	6	2
Probability	0.000	0.027	0.024	0.005	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Correlation Matrix											
$y^{(3m)}$	1	0.983	0.966	0.950	0.937	0.925	0.915	0.905	0.896	0.887	0.878
$y^{(1y)}$		1	0.993	0.982	0.970	0.959	0.948	0.938	0.929	0.919	0.910
$y^{(2y)}$			1	0.997	0.990	0.983	0.975	0.967	0.958	0.950	0.942
$y^{(3y)}$				1	0.998	0.994	0.989	0.983	0.976	0.970	0.963
$y^{(4y)}$					1	0.999	0.996	0.992	0.987	0.982	0.976
$y^{(5y)}$						1	0.999	0.997	0.994	0.990	0.985
$y^{(6y)}$							1	0.999	0.997	0.995	0.991
$y^{(7y)}$								1	0.999	0.998	0.996
$y^{(8y)}$									1	1.000	0.998
$y^{(9y)}$										1	1.000
$y^{(10y)}$											1

We report the estimates of parameters for the aforementioned models in Table IV-3. Since the unconstrained nature of regimes 0 and 1 in estimation, we need to identify the regimes *ex post*. We find that, for all models, the variances of the error term in short rate equation and those of the 5-year and 10-year yields are higher in regime 0 than in regime 1,

that is $\sigma_0 > \sigma_1$, $\Sigma_0^{(5)} > \Sigma_1^{(5)}$ and $\Sigma_0^{(10)} > \Sigma_1^{(10)}$. Meanwhile, the speed of mean-reversion is higher in regime 0 than in regime 1, that is $k_0 > k_1$. In addition, the long term mean of the instantaneous short rate is higher in regime 0 than in regime 1 (except model 1 and 4). This suggests that regime 0 can be identified as the high volatility regime for the short rate with higher speed of adjustment to its mean value, and the short rate in regime 1 incorporates a lower volatility with a lower speed of mean reversion. This finding is consistent with previous studies on short rates, which report that the short rate tends to revert much quicker to its mean when it is further away from its long term mean value or during volatile time periods.

Table IV-3: Parameter estimates of various models

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
k_0	1.114E-02 (1.297E-03)	1.002E-02 (2.668E-03)	1.471E-14 (5.556E-05)	9.838E-03 (2.441E-03)	1.631E-02 (2.282E-03)	3.499E-13 (1.366E-04)	1.694E-14 (1.262E-04)	1.770E-02 (2.481E-03)
k_1		2.043E-03 (9.768E-04)		6.022E-03 (2.096E-03)	9.962E-14 (2.139E-04)			3.017E-12 (2.806E-04)
p		8.932E-01 (1.635E-02)	9.106E-01 (1.836E-02)	9.001E-01 (1.501E-02)	9.295E-01 (9.840E-03)	9.051E-01 (1.820E-02)	9.408E-01 (1.335E-02)	9.272E-01 (1.011E-02)
q		9.597E-01 (9.637E-03)	9.599E-01 (1.116E-02)	9.508E-01 (9.861E-03)	9.662E-01 (4.749E-03)	9.482E-01 (1.283E-02)	9.742E-01 (9.031E-03)	9.592E-01 (5.583E-03)
δ_0	8.746E-03 (4.012E-04)	8.701E-10 (3.352E-04)	2.816E-01 (70.680)	1.073E-09 (2.314E-04)	1.775E-09 (3.427E-04)	7.224E-03 (36.989)	4.709E-02 (277.042)	1.030E-09 (2.895E-04)
δ_1		3.640E-03 (4.582E-04)	8.026E-04 (57.801)		1.705E-03 (63.112)		1.465E-03 (32.958)	
σ_0	5.555E-03 (1.483E-04)	7.540E-03 (3.296E-04)	7.453E-03 (3.193E-04)	7.493E-03 (3.163E-04)	5.805E-03 (1.787E-04)	7.317E-03 (3.091E-04)	5.517E-03 (1.528E-04)	5.796E-03 (1.785E-04)
σ_1		1.967E-03 (1.176E-04)	1.739E-03 (1.054E-04)	1.993E-03 (1.216E-04)		2.012E-03 (1.600E-04)		
λ_0	-1.377E-02 (1.274E-03)	-5.178E-12 (2.846E-04)	-9.698E-14 (1.623E-04)	-1.739E-12 (2.222E-04)	-2.212E-12 (5.214E-04)	-5.423E-15 (1.765E-04)	-1.972E-13 (2.030E-04)	-2.604E-12 (6.042E-04)
λ_1		-2.615E-02 (4.368E-04)	-3.121E-03 (5.032E-04)	-1.302E-02 (1.917E-03)	-1.023E-02 (1.107E-03)	-3.066E-03 (5.542E-04)	-3.645E-03 (4.498E-04)	-1.202E-02 (1.381E-03)
$\Sigma_0^{(5)}$	1.207E-03 (3.279E-05)	1.669E-03 (8.677E-05)	1.684E-03 (8.906E-05)	1.652E-03 (8.379E-05)	1.456E-03 (9.357E-05)	1.720E-03 (9.235E-05)	1.778E-03 (1.025E-04)	1.462E-03 (1.019E-04)
$\Sigma_0^{(10)}$	1.489E-03 (4.056E-05)	2.152E-03 (1.120E-04)	2.057E-03 (1.093E-04)	2.129E-03 (1.079E-04)	2.060E-03 (1.325E-04)	2.159E-03 (1.163E-04)	2.260E-03 (1.299E-04)	2.068E-03 (1.438E-04)
$\rho_0^{(5,10)}$	9.250E-01 (5.881E-03)	9.659E-01 (4.899E-03)	9.666E-01 (5.007E-03)	9.646E-01 (4.988E-03)	9.730E-01 (4.706E-03)	9.637E-01 (5.370E-03)	9.632E-01 (5.637E-03)	9.738E-01 (4.850E-03)
$\Sigma_1^{(5)}$		7.283E-04 (1.356E-05)	8.197E-04 (1.925E-05)	7.212E-04 (2.641E-05)	9.157E-04 (1.957E-05)	7.085E-04 (2.576E-05)	8.235E-04 (2.444E-05)	9.100E-04 (2.096E-05)
$\Sigma_1^{(10)}$		9.110E-04 (7.875E-06)	1.176E-03 (3.159E-05)	9.058E-04 (3.739E-05)	9.878E-04 (2.185E-05)	8.988E-04 (3.338E-05)	9.745E-04 (2.687E-05)	9.839E-04 (2.308E-05)
$\rho_1^{(5,10)}$		9.690E-01 (4.405E-03)	9.768E-01 (3.239E-03)	9.697E-01 (3.724E-03)	9.749E-01 (2.822E-03)	9.688E-01 (5.462E-03)	9.770E-01 (2.503E-03)	9.746E-01 (2.956E-03)
$\ln L$	8567.227	8921.011	8898.793	8917.548	8824.506	8900.791	8780.124	8824.129
AIC	-35.083	-36.496	-36.409	-36.486	-36.105	-36.421	-35.927	-36.107
BIC	-35.023	-36.359	-36.280	-36.357	-35.976	-36.301	-35.807	-35.987

Note: The standard errors, which are shown in parentheses, are numerically calculated using finite difference method.

In Figure IV-2, we plot the smoothed state probabilities of the short rate being in regime 0 for each model. Although the smoothed state probabilities for each model differ slightly from each other, the general pattern we see is that the short rate tends to enter regime 0 (a volatile regime with higher speed of mean reversion) when the yield spreads are wider (see the comparison with the evolution of nominal yields in Figure IV-1). On the other hand, the short rate tends to stay in regime 1 (a low volatility regime with lower speed of mean reversion) when yields with different maturities co-move closely with each other. Since cross-sectional co-movements of all the yields means smaller bond pricing errors while we use a same set of parameters to price bonds with different maturities, the resulting bond pricing errors should be less volatile in regime 1. Theoretically, a tightening yield spread under less volatile market conditions (regime 1) means a flatter yield curve, and this is generally associated with the expansion phase of the economy. Conversely, if the yield spreads are widening (e.g. the recent financial crisis which resulted a widening spread), the movements of bond prices for different maturities are diverging with each other, leading to an increase in bond pricing errors (as in regime 0).

Figure IV-1: UK zero-coupon bond yields (% , January 1970 to September 2010)

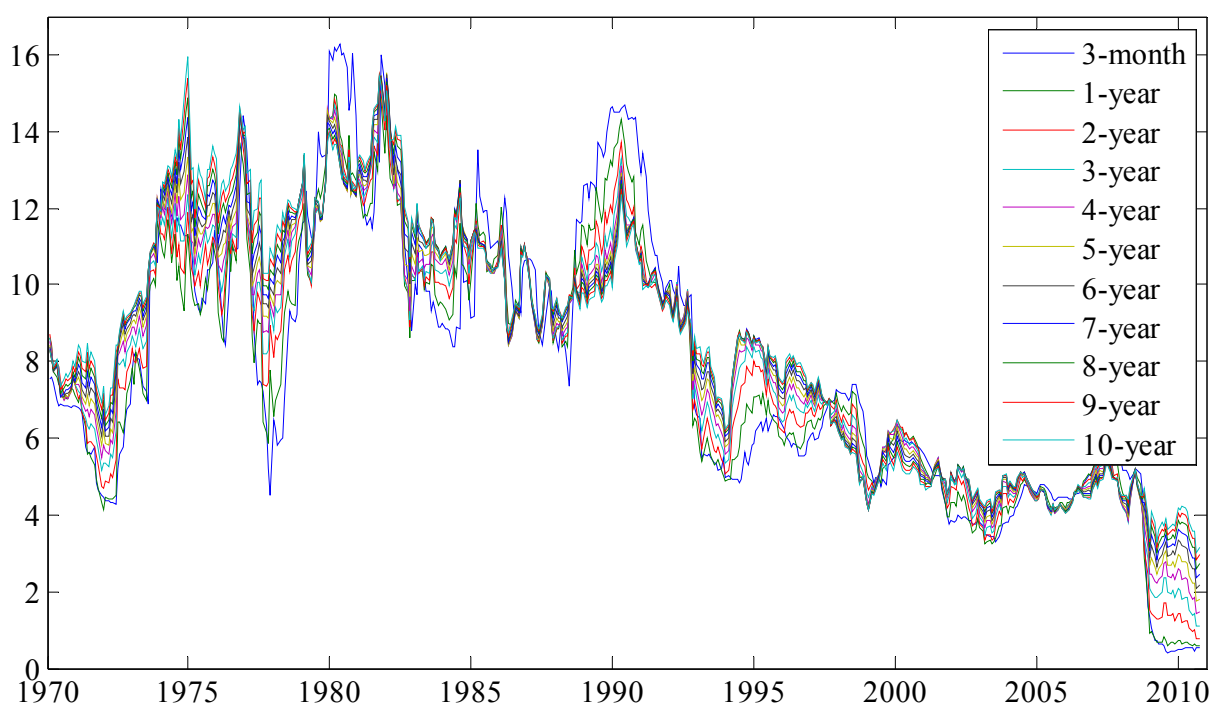
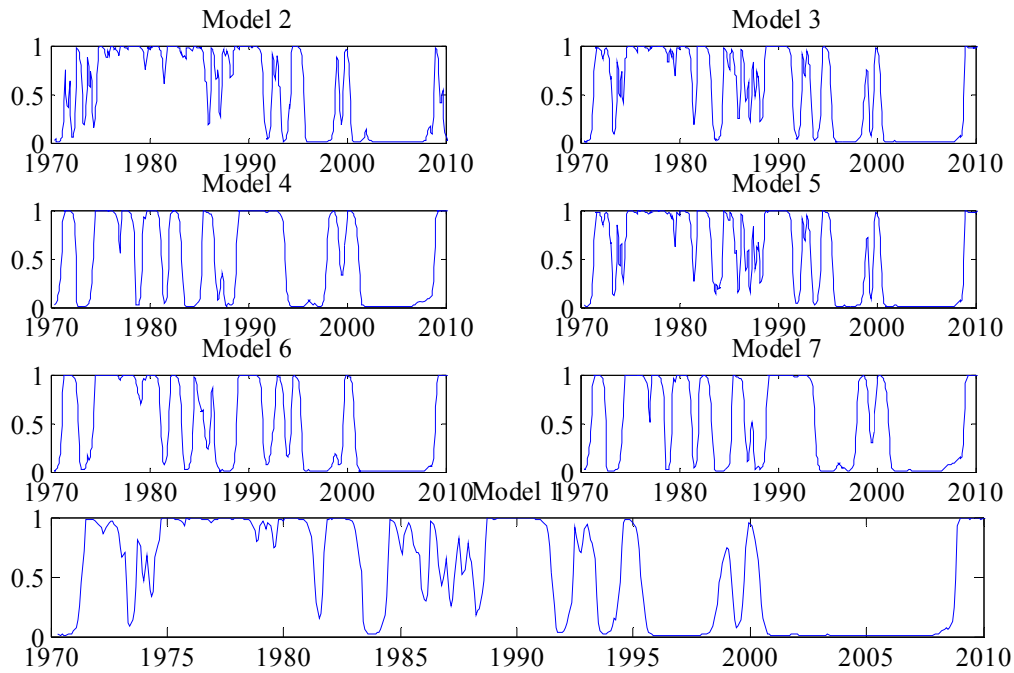


Figure IV-2: Smoothed state probabilities of regime 0 for various models



Having estimated one constant parameter CIR model and 7 Markov switching CIR models, finding the best in-sample performing model is a cumbersome task. Of the comparison between estimated yields and observed ones, a visual inspection (Figure IV-3 to Figure IV-5) is difficult to tell which model has the best in-sample performance. The standard information selection criteria - AIC and BIC, which are reported in the last two rows of Table IV-3, suggest Model 1 as the best in-sample performing mode. This is not surprised because Model 1, comparing with the rest models which are nested to Model 1, is fitted with more parameters and hence the most flexible model amongst the eight. The conventional likelihood ratio test is the most widely used tests applied on nested models. We present the likelihood ratio test results in Table IV-4, where the simplified (or restricted models) are displayed horizontally and the comparatively more flexible models are ordered vertically. For example, Model 1 is the most flexible (least restricted) model allowing all parameters to switch between regimes, while Models 2 to 7 are comparatively restricted models with one or two parameters constrained to be invariant over time. Since Models 1 nests Models 2 to 7, we are able to test the restrictions applied on Models 2 to 7 using likelihood ratio tests. Similarly, we can also test the additional restriction that δ is a constant (Model 5) against Model 2 (where k is assumed to be a constant).

Figure IV-3: Observed 3-month rates and fitted 3-month rates from models

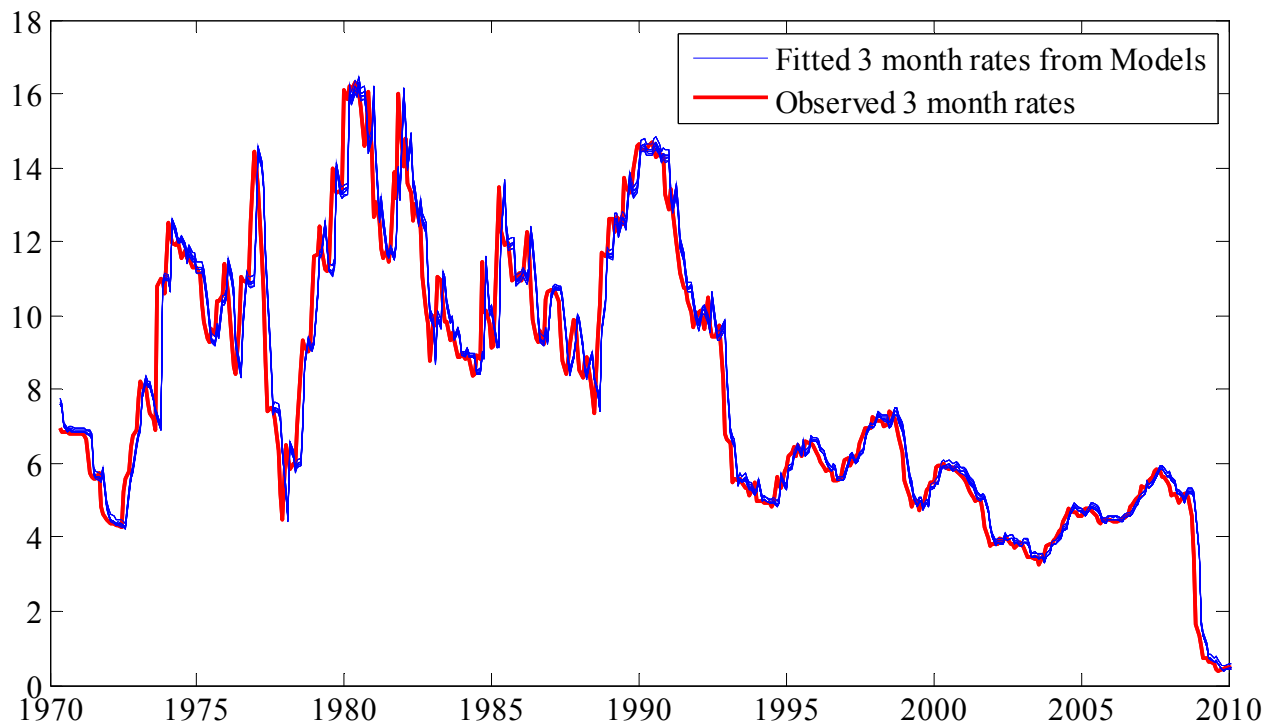


Figure IV-4: Observed 5-year rates and fitted 5-year rates from models

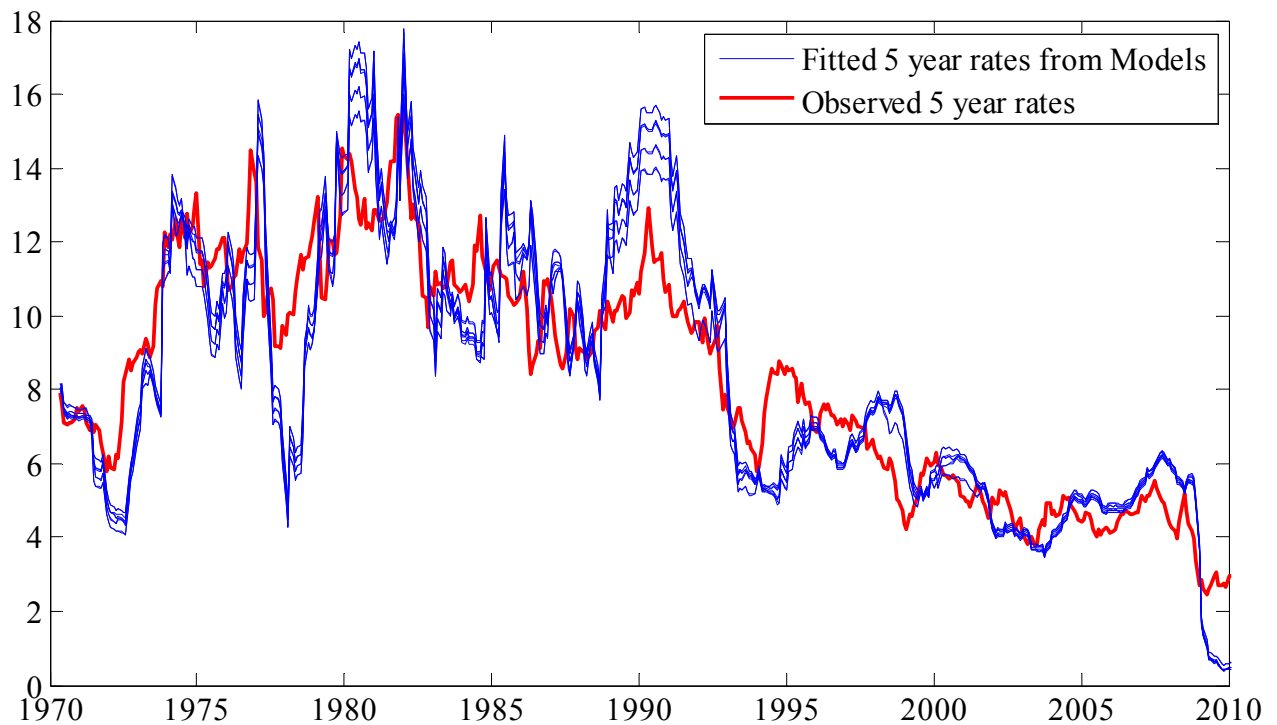


Figure IV-5: Observed 10-year rates and fitted 10-year rates from models

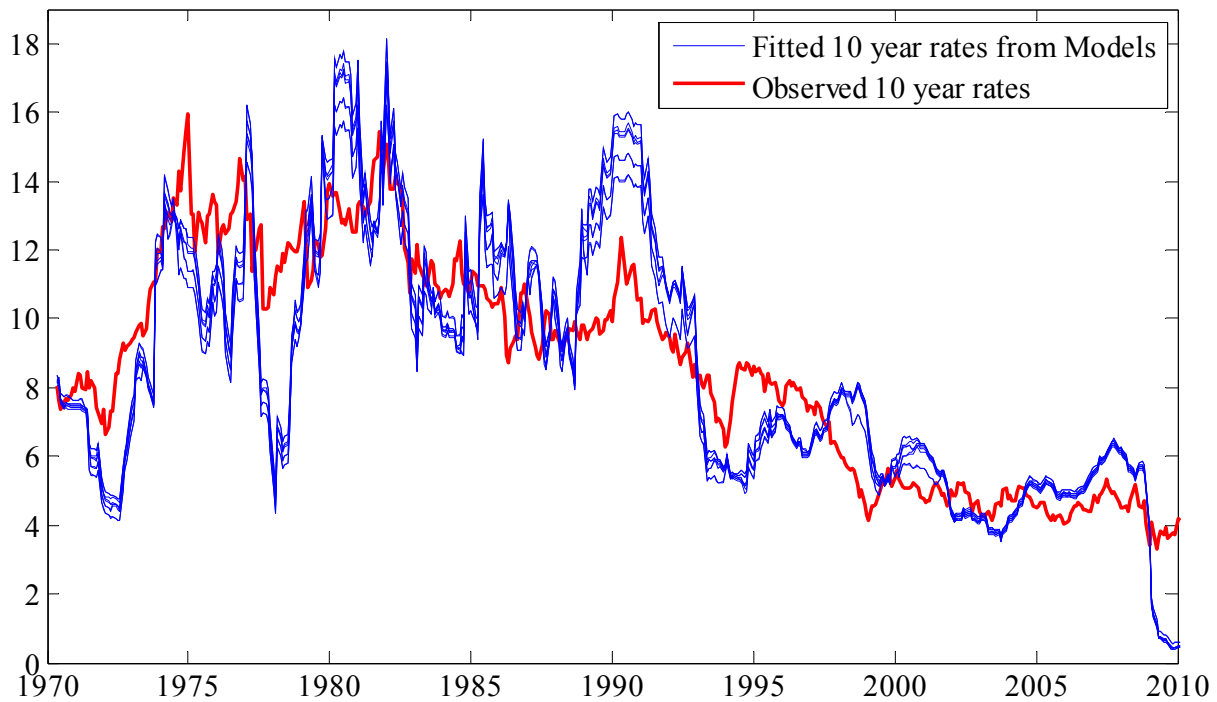


Table IV-4: Likelihood ratio tests¹⁰

	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Model 1	44.437 [2.627E-11]	6.926 [8.497E-03]	193.010 [7.004E-44]	40.440 [1.654E-09]	281.775 [6.508E-62]	193.763 [8.411E-43]
Model 2	-	-	-	3.997 [0.046]	237.337 [1.497E-53]	-
Model 3	-	-	-	33.514 [7.075E-09]	-	186.838 [1.558E-42]
Model 4	-	-	-	-	88.764 [4.447E-21]	0.753 [0.385]

Note: Chi-square distributions of the likelihood ratio statistics are shown in square brackets.

The likelihood ratio tests shown above indicate that all alternative simplified models (Model 2-7) are rejected against the most flexible model. The hypothesis that δ is a constant (Model 5) against Model 2 where δ is assumed to be regime dependent, is weakly rejected at 5% significance level. However, the same null hypothesis applied on Model 7 against the alternative in Model 4 is statistically and significantly accepted. This suggests that assuming a regime dependent long term mean of the short rate is of secondary importance in practice, and

¹⁰ We do not report the likelihood ratio statistics on the most restricted models against the rest of the less restricted models. It is well known that for Markov-switching models the standard likelihood ratio test of the null hypothesis of linearity does not have the usual χ^2 distribution. The reason is that there are nuisance parameters which cannot be identified under the null hypothesis. As a result, the scores evaluated at the null hypothesis are identically zero. Hansen (1992) and Garcia (1998) introduce alternative tests of the linearity against regime switching. In this paper, we use Hansen (1992) procedure, which provides an upper bound of the p-value for linearity, to determine the significance of improvement of allowing Markov-switching disturbance terms in the two components.

the problem of over parameterization may cause flawed interpretation on the dynamics of the short rate.

Since our sample period begins from the time when the UK economy was recovering from the second oil crisis, and ends at the time when the recent financial crisis brought the short rate to a near zero level (see Figure IV-6), the estimation results may be sensitive to the choice of sample periods. In order to see whether this is the case we follow Driffill, *et al.* (2009) to calculate recursive AIC and BIC information criteria. Starting from the period between January 1970 and January 2003, we recursively estimate all the aforementioned models by sequentially enlarge the sample up to September 2010. The AIC and BIC information criteria are then recursively computed for each sample.

Figure IV-6: Snap shot of the yields between January 2003 and September 2010

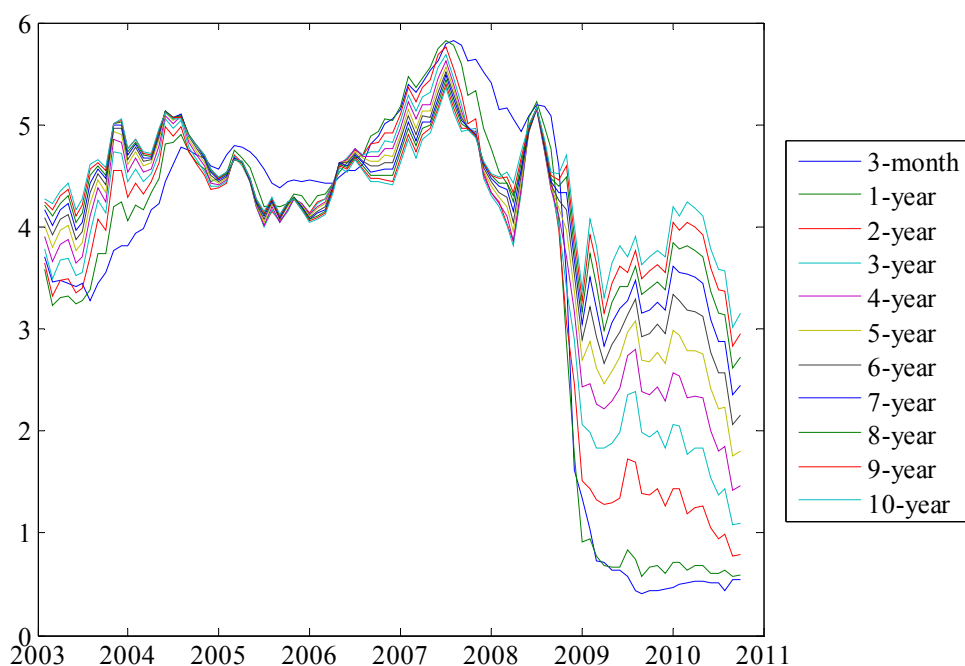


Figure IV-7: Recursive AIC and BIC

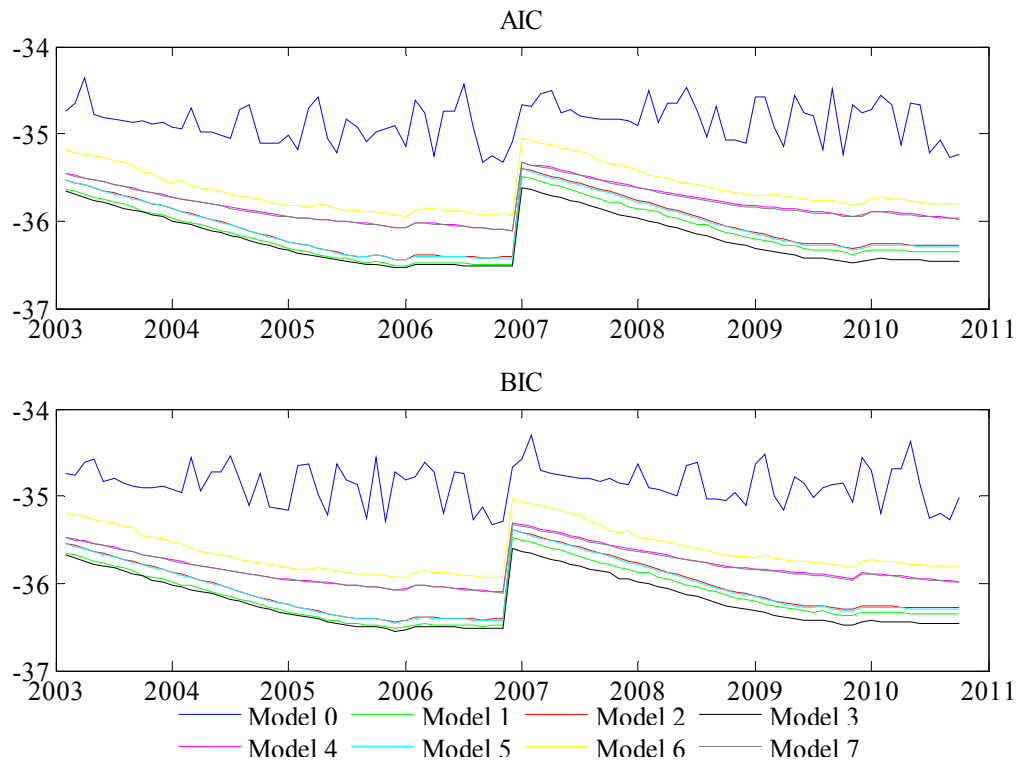


Figure IV-7 plots the recursive AIC and BIC information criteria over the period from January 2003 to September 2010. It is clearly depicted that Model 0 is not preferred over the whole sample period on the basis of both AIC and BIC information criteria. Interestingly, Model 3 seems to be preferred against all other models, although the difference between it and Model 1 is almost indistinguishable before 2007. One may have noticed the spike of the two information criteria for all models at the end of 2006. More specifically, the model which assumes different long term average of the short rate (Model 3) is much clearly shown to be preferred to the most flexible model (Model 1) after the spike. One possible explanation of the spike is the fact that we used the same set of initial values in the recursive AIC and BIC calculations. If the initial values, which are derived from the estimation on data set ranging from January 1970 to January 2003, are sensitive to the changes in sample periods (i.e. when we enlarge the sample size at each time point from February 2003 to September 2010), the likelihood function may not be globally maximized, hence the spike of the two information criteria. This finding echoes with our previous doubts that model specifications are sensitive to the sample specification. To circumvent this problem, one can use the parameter estimates returned from previous calculation as the initial values for the next calculation. Using updated initial values in the recursive calculation implicitly assumes that the parameters are time-varying and hence improves the fitness of models. This is, in sense, equivalent to the forecasting exercise we shall do in section IV.5.

IV.5 Ex-ante pricing performance of the models

The evaluations of each model's goodness-of-fit in the previous section are based on in-sample estimation. We find, however, conflicting results. In this section, we follow Driffill, *et al.* (2009) to assess the models' *ex-ante* bond pricing performance. We assume that, at time t , the coefficients $A_{n+1}(i)$ and $B_{n+1}(i)$ in bond pricing equation are obtained using estimates at time $t-1$, which we denote $A_{n+1,t-1}(S_{t-1})$ and $B_{n+1,t-1}(S_{t-1})$. Since we know the short rate, which we assume is exogenous, at time t is x_t , a recursive bond pricing exercise (forecasting the long yields with maturities 1, 2, 3, 4, 6, 7, 8 and 9 years using information at time $t-1$ and x_t) can be conducted while we sequentially enlarge the sample period from January 1970-January 2003 to January 1970-September 2010¹¹. This recursive forecasting exercise gives us a total of 92 sample points. We compare these forecasted yields with the observed yields for the sample period January 2003 to September 2010 by calculating the standard Mean Square Error (MSE) and the Forecasting Direction- Test(FDT) with the Pesaran-Timmermann statistics (PT) proposed by Pesaran and Timmermann (1992) and Pesaran and Timmermann (1994).

PT test is a generalization of the Henriksson-Merton test for independence between forecast and actual values (see Henriksson and Merton (1981)). It examines whether the direction of the movements of the forecast and actual values are in match with each other. The larger the Pesaran-Timmermann statistics, the better they are in step with each other. It, however, does not take into account the magnitude of the difference between the forecast and actual values, which is why it is known as a forecasting directional test.

Let's denote Δy_t as the actual change of the yields for the period January 2003 to September 2010. The forecast counterpart is denoted as $\Delta \hat{y}_t$. Now, we define

$$X_t \begin{cases} = 1 & \text{if } \Delta y_t > 0 \\ = 0 & \text{if } \Delta y_t \leq 0 \end{cases}, Y_t \begin{cases} = 1 & \text{if } \Delta \hat{y}_t > 0 \\ = 0 & \text{if } \Delta \hat{y}_t \leq 0 \end{cases}, Z_t \begin{cases} = 1 & \text{if } \Delta y_t \Delta \hat{y}_t > 0 \\ = 0 & \text{if } \Delta y_t \Delta \hat{y}_t \leq 0 \end{cases}.$$

The FDT statistics, as shown in Table IV-5, are calculated as one minus the summation of Z_t divided by the total number of forecasted points (in our case, there are 92 points). Since Z_t corresponds to the correct prediction of direction, the division gives us the proportion of correctly predicted directional movements of the forecasts. One minus this number gives us the proportion of wrong prediction, which is also called the confusion rate. Generally speaking, the

¹¹ This exercise is equivalent to the situation that we want to forecast the yields with maturities 1,2,3,4,6,7,8 and 9 years at time t but only have the information on yields with maturities 3-month, 5 and 10 years at time $t-1$ and the short rate at time t .

FDT statistics indicate that all models perform better at predicting directional movements of the yields with shorter maturities. The longer the maturities, the higher proportion those models predict wrongly the directional movements of yields. Overall, the best performing model is Model 3 with a FDT statistic of 28.7%, which is lower than Model 1's 29.4%. However, in terms of MSE measure, which takes into account the magnitude of the forecast errors, Model 1 performs slightly better than Model 3.

Table IV-5: MSE and FDT statistics

	1-year	2-year	3-year	4-year	6-year	7-year	8-year	9-year	Total
MSE (%)									
Model 0	0.086	0.322	0.639	0.978	1.651	1.976	2.284	2.572	10.508
Model 1	0.060	0.270	0.581	0.918	1.585	1.906	2.212	2.498	10.031
Model 2	0.068	0.293	0.607	0.945	1.615	1.939	2.247	2.535	10.248
Model 3	0.060	0.270	0.583	0.921	1.590	1.912	2.219	2.506	10.061
Model 4	0.071	0.291	0.618	0.964	1.639	1.963	2.271	2.561	10.378
Model 5	0.068	0.293	0.607	0.944	1.615	1.939	2.248	2.536	10.250
Model 6	0.067	0.292	0.606	0.944	1.618	1.943	2.253	2.543	10.266
Model 7	0.071	0.292	0.618	0.964	1.639	1.963	2.272	2.561	10.380
FDT									
Model 0	0.286	0.352	0.385	0.363	0.363	0.374	0.363	0.374	0.357
Model 1	0.220	0.253	0.308	0.297	0.319	0.330	0.308	0.319	0.294
Model 2	0.231	0.264	0.308	0.286	0.308	0.319	0.297	0.308	0.290
Model 3	0.231	0.253	0.297	0.286	0.308	0.319	0.297	0.308	0.287
Model 4	0.242	0.275	0.319	0.297	0.319	0.330	0.308	0.319	0.301
Model 5	0.231	0.264	0.308	0.286	0.308	0.319	0.297	0.308	0.290
Model 6	0.231	0.264	0.308	0.286	0.308	0.319	0.297	0.308	0.290
Model 7	0.242	0.275	0.319	0.297	0.319	0.330	0.308	0.319	0.301

Now, we produce the Pesaran-Timmermann statistic, which is defined as

$$PT = \frac{P - P^*}{\sqrt{\text{Var}(P) - \text{Var}(P^*)}} \xrightarrow{d} N(0,1),$$

where

$$P = \frac{1}{T} \sum_{t=1}^T Z_t, \quad P_X = \frac{1}{T} \sum_{t=1}^T X_t, \quad P_Y = \frac{1}{T} \sum_{t=1}^T Y_t,$$

and

$$P^* = P_X P_Y + (1 - P_X)(1 - P_Y).$$

Note the variances of P and P^* are calculated as

$$Var(P) = \frac{1}{T} P^* (1 - P^*)$$

$$Var(P^*) = \frac{1}{T} P_x (1 - P_x) (2P_y - 1)^2 + \frac{1}{T} P_y (1 - P_y) (2P_x - 1)^2 + \frac{4}{T^2} P_x P_y (1 - P_x) (1 - P_y).$$

Table IV-6: Pesaran-Timmermann test

	1-year	2-year	3-year	4-year	6-year	7-year	8-year	9-year
Model 0	3.878 [5.276E-05]	2.836 [2.287E-03]	2.631 [4.262E-03]	2.223 [1.311E-02]	2.239 [1.256E-02]	2.223 [1.311E-02]	2.456 [7.034E-03]	2.503 [6.152E-03]
Model 1	4.712 [1.228E-06]	4.330 [7.461E-06]	3.766 [8.307E-05]	3.164 [7.773E-04]	3.224 [6.331E-04]	3.164 [7.773E-04]	3.588 [1.668E-04]	3.588 [1.668E-04]
Model 2	4.922 [4.285E-07]	4.094 [2.118E-05]	4.114 [1.944E-05]	3.299 [4.856E-04]	3.331 [4.334E-04]	3.299 [4.856E-04]	3.720 [9.947E-05]	3.720 [9.947E-05]
Model 3	4.920 [4.327E-07]	4.310 [8.144E-06]	4.340 [7.128E-06]	3.299 [4.856E-04]	3.331 [4.334E-04]	3.299 [4.856E-04]	3.720 [9.947E-05]	3.720 [9.947E-05]
Model 4	4.710 [1.241E-06]	3.891 [5.001E-05]	3.919 [4.441E-05]	3.112 [9.298E-04]	3.152 [8.109E-04]	3.112 [9.298E-04]	3.534 [2.048E-04]	3.534 [2.048E-04]
Model 5	4.922 [4.285E-07]	4.094 [2.118E-05]	4.114 [1.944E-05]	3.299 [4.856E-04]	3.331 [4.334E-04]	3.299 [4.856E-04]	3.720 [9.947E-05]	3.720 [9.947E-05]
Model 6	4.922 [4.285E-07]	4.094 [2.118E-05]	4.114 [1.944E-05]	3.299 [4.856E-04]	3.331 [4.334E-04]	3.299 [4.856E-04]	3.720 [9.947E-05]	3.720 [9.947E-05]
Model 7	4.710 [1.241E-06]	3.891 [5.001E-05]	3.919 [4.441E-05]	3.112 [9.298E-04]	3.152 [8.109E-04]	3.112 [9.298E-04]	3.534 [2.048E-04]	3.534 [2.048E-04]

Note: The asymptotic p-values are shown in square brackets.

The Pesaran-Timmermann statistic, under the null hypothesis that Δy_t and $\Delta \hat{y}_t$ are independent random variables, follows an asymptotic standard normal distribution. The statistics calculated for each model are reported in Table IV-6. We find that Model 2 outperforms other competing models in forecasting the directional movements of yields with maturity of 1 year, while Model 1 and Model 3 are better performing models in forecasting the yields with maturities of 2-year and 3-year, respectively. For the rest of the yields, Models 2, 3, 5 and 6 are equally achieving the highest Pesaran-Timmermann statistic, which suggests that they are all better than the most flexible model in predicting the directional movements of yields with maturities 4, 6, 7, 8 and 9 years.

IV.6 Conclusion

In this study, we apply the methodology used in Driffill, *et al.* (2009) to the UK's term structure of interest rates. We conduct a series of model specification tests on a battery of models that apply different parameter restrictions on the short rate equation. We also evaluate the out-of-sample forecasting performance of each model using the forecasting direction test and the Pesaran-Timmermann statistics. In a Markov switching CIR model framework, our results indicate that the least restricted model specification (Model 1) gives the best goodness-of-fit on the UK's term structure of interest rates. Although other simplified models (especially

Model 3) may provide slightly better out-of-sample forecast in directional movements of the yields, Model 1 still achieves the lowest MSE. In other words, this suggests that the economic gains of employing simplified Markov switching CIR models on the UK term structure of interest rates is relatively small.

Bibliography

- Bansal, Ravi, and Hao Zhou, 2002, Term structure of interest rates with regime shifts, *Journal of Finance* 57, 1997-2043.
- Cai, Jun, 1994, A markov model of switching-regime arch, *Journal of Business & Economic Statistics* 12, 309-317.
- Driffill, John, Turalay Kenc, Martin Sola, and Fabio Spagnolo, 2009, The effects of different parameterizations of markov-switching in a cir model of bond pricing, *Studies in Nonlinear Dynamics & Econometrics* 13.
- Garcia, R., 1998, Asymptotic null distribution of the likelihood ratio test in markov switching models, *International Economic Review* 39, 763-788.
- Gray, Stephen F., 1996, Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics* 42, 27-62.
- Hamilton, James D. , 1989b, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica: Journal of the Econometric Society* 57, 357-384.
- Hansen, B. E., 1992, The likelihood ratio test under nonstandard conditions: Testing the markov switching model of gnp, *Journal of Applied Econometrics* 7, S61-S82.
- Henriksson, Roy D. , and Robert C. Merton, 1981, On market timing and investment performance. Ii. Statistical procedures for evaluating forecasting skills, *Journal of Business* 54, 513-533.
- Naik, Vasant, and Moon Hoe Lee, 1997, Yield curve dynamics with discrete shifts in economic regimes: Theory and estimation, *Working Paper* Faculty of Commerce, University of British Columbia.
- Pesaran, M. H., and A. G. Timmermann, 1994, A generalization of the non-parametric henriksson-merton test of market timing, *Economics Letters* 44, 1-7.
- Pesaran, M. Hashem, and Allan Timmermann, 1992, A simple nonparametric test of predictive performance, *Journal of Business & Economic Statistics* 10, 461-466.

Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177-188.

Appendix IV

Matlab code used for Markov switching CIR model

The following codes require CompEcon Toolbox to run. The toolbox can be downloaded from www4.ncsu.edu/~pfackler/compecon

```
%% =====run_uk.m=====
%% This is the main code file, it calls the likhfcn.m to minimize the
%% negative value of the likelihood value, calls the hfilter.m to
%% filter out the switching probabilities
%% =====

clear all
close all
global sr ly t longest_month_maturity
%% LOAD DATA
% 3 month short rate from Jan 1970 to Feb 1997 is obtained from Datastream
% 1 year to 10 year rates from Jan 1970 to Sep 2010 are all from Bank of
% England
load UKNom3M_10Y_Jan70_Sep10.txt;
date_t = UKNom3M_10Y_Jan70_Sep10(:,1); % Jan 1970 to Sep 2010
data = UKNom3M_10Y_Jan70_Sep10(:,2:12)/1200;
total_t = size(data,1); % 1 to 489

%% CONTROL VARIABLES
recursive_estimation=1; % 1 if model fit test requires recursive estimation
freq=12; % monthly data, so the frequency of observation is 12 for a year
longest_year_maturity=10; % the longest year of maturity is 10 years
longest_month_maturity=longest_year_maturity*freq; % the longest month of maturity

recursive_start=397; % start from Jan 2003

%% PREPARE DATA
if recursive_estimation==0 % then full sample estimation
    sr=data(:,1);
    ly=data(:,2:11);
    t=size(sr,1);
    para_in=[0.00718495910436238;-0.00542577700971817;-0.520703719428048;-
0.509132004828306;8.68769923139257e-10;-
0.00155070566024247;0.00756517659770299;0.00174021287872868;-1.35236631329832e-12;-
0.0112375002689871;-0.00163808554163196;0.00202676104113847;28.3136852542254;-
0.000837523314938341;-0.00119746535266108;47.1583644468502];
    % optimization
    options=optimset('Display','iter','TolX',1e-4,'TolFun',1e-8,...
'MaxFunEvals',1000000,'MaxIter',1000000);
    [xout,fval,exitflag,output]=fminsearch(@likhfcn,para_in,options);
    %[xout,fval,exitflag,output]=fminunc(@likhfcn,para_in,options);
    % estimation results
    xfnl=trans_para(xout);
    % S.E. and t-test
    h_0=fdhess(@likhfcn,xout); % calculate Hessian matrix
    g_0=fdjac1(@trans_para,xout,[]); % calculate gradient
    h_fnl=g_0*(inv(h_0))*g_0'; % final Hessian
    std_fnl=sqrt(diag(h_fnl)); % S.E.
    t_ratio=xfnl/std_fnl; % t-ratio
    % print results
    table_full_sample_para=[xfnl std_fnl t_ratio];
    sprintf('=====')
    sprintf('Est.Para.~Std.Errors.~t-ratio')
```

```

disp(table_full_sample_para)
% save results
fid=fopen('full_sample_result.txt','wt');
fprintf(fid,'k0 %8.3f %8.3f %8.3f\n',table_full_sample_para(1,:));
fprintf(fid,'k1 %8.3f %8.3f %8.3f\n',table_full_sample_para(2,:));
fprintf(fid,'p %8.3f %8.3f %8.3f\n',table_full_sample_para(3,:));
fprintf(fid,'q %8.3f %8.3f %8.3f\n',table_full_sample_para(4,:));
fprintf(fid,'delta0 %8.3f %8.3f %8.3f\n',table_full_sample_para(5,:));
fprintf(fid,'delta1 %8.3f %8.3f %8.3f\n',table_full_sample_para(6,:));
fprintf(fid,'sig0 %8.3f %8.3f %8.3f\n',table_full_sample_para(7,:));
fprintf(fid,'sig1 %8.3f %8.3f %8.3f\n',table_full_sample_para(8,:));
fprintf(fid,'lam0 %8.3f %8.3f %8.3f\n',table_full_sample_para(9,:));
fprintf(fid,'lam1 %8.3f %8.3f %8.3f\n',table_full_sample_para(10,:));
fprintf(fid,'siguu0 %8.3f %8.3f %8.3f\n',table_full_sample_para(11,:));
fprintf(fid,'sigud0 %8.3f %8.3f %8.3f\n',table_full_sample_para(12,:));
fprintf(fid,'corr0 %8.3f %8.3f %8.3f\n',table_full_sample_para(13,:));
fprintf(fid,'siguu1 %8.3f %8.3f %8.3f\n',table_full_sample_para(14,:));
fprintf(fid,'sigud1 %8.3f %8.3f %8.3f\n',table_full_sample_para(15,:));
fprintf(fid,'corr1 %8.3f %8.3f %8.3f\n',table_full_sample_para(16,:));
fprintf(fid,'loglik %8.3f\n',fval);
fclose(fid);

elseif recursive_estimation==1 % then
counter=1;
for j_iter=recursive_start:total_t
sr=data(1:j_iter,1);
ly=data(1:j_iter,2:11);
t=size(sr,1);
para_in=[0.00718495910436238;-0.005425777700971817;-0.520703719428048;-
0.509132004828306;8.68769923139257e-10;-
0.00155070566024247;0.00756517659770299;0.00174021287872868;-1.35236631329832e-12;-
0.0112375002689871;-0.00163808554163196;0.00202676104113847;28.3136852542254;-
0.000837523314938341;-0.00119746535266108;47.1583644468502];
% optimization
options=optimset('Display','iter','TolX',1e-4,'TolFun',1e-8,...
'MaxFunEvals',1000000,'MaxIter',100000);
[xout,fval,exitflag,output]=fminsearch(@likhfcn,para_in,options);
% estimation results
xfnl=trans_para(xout);
% S.E. and t-test
h_0=fdhess(@likhfcn,xout); % calculate Hessian matrix
g_0=fdjac1(@trans_para,xout,[]); % calculate gradient
h_fnl=g_0*(inv(h_0))*g_0'; % final Hessian
std_fnl=sqrt(diag(h_fnl)); % S.E.
t_ratio=xfnl./std_fnl; % t-ratio
% print results
table_full_sample_para=[xfnl std_fnl t_ratio];
%sprintf('=====')
%sprintf('Est.Para.~Std.Errors.~t-ratio')
%disp(table_full_sample_para)
% save results
counter_name=sprintf('recursive_%d.txt',counter);
fid=fopen(counter_name,'wt');
fprintf(fid,'k0 %8.3f %8.3f %8.3f\n',table_full_sample_para(1,:));
fprintf(fid,'k1 %8.3f %8.3f %8.3f\n',table_full_sample_para(2,:));
fprintf(fid,'p %8.3f %8.3f %8.3f\n',table_full_sample_para(3,:));
fprintf(fid,'q %8.3f %8.3f %8.3f\n',table_full_sample_para(4,:));
fprintf(fid,'delta0 %8.3f %8.3f %8.3f\n',table_full_sample_para(5,:));
fprintf(fid,'delta1 %8.3f %8.3f %8.3f\n',table_full_sample_para(6,:));
fprintf(fid,'sig0 %8.3f %8.3f %8.3f\n',table_full_sample_para(7,:));

```

```

fprintf(fid,'sig0 %8.3f%8.3f%8.3f\n',table_full_sample_para(8,:));
fprintf(fid,'lam0 %8.3f%8.3f%8.3f\n',table_full_sample_para(9,:));
fprintf(fid,'lam1 %8.3f%8.3f%8.3f\n',table_full_sample_para(10,:));
fprintf(fid,'siguu0 %8.3f%8.3f%8.3f\n',table_full_sample_para(11,:));
fprintf(fid,'sigud0 %8.3f%8.3f%8.3f\n',table_full_sample_para(12,:));
fprintf(fid,'corru0 %8.3f%8.3f%8.3f\n',table_full_sample_para(13,:));
fprintf(fid,'siguu1 %8.3f%8.3f%8.3f\n',table_full_sample_para(14,:));
fprintf(fid,'sigud1 %8.3f%8.3f%8.3f\n',table_full_sample_para(15,:));
fprintf(fid,'corrud1 %8.3f%8.3f%8.3f\n',table_full_sample_para(16,:));
fprintf(fid,'loglik %8.3f\n',-fval);
fprintf(fid,'No. Observations is %8.3f\n',t);
fclose(fid);
counter=1+counter;
end
end

```

%% filter

```

[stl, stt, fit_5y, fit_10y, fit_3m]=hfilter(xout);

```

```

function minus_lik_value=likhfcn(para_in)
% This function evaluate the likelihood function of the markov switching CIR
global sr ly t longest_month_maturity
para=trans_para(para_in);

```

% assign parameters

```

k0=para(1);

```

```

k1=para(2);

```

```

p=para(3);

```

```

q=para(4);

```

```

ssp0=(1-q)/(2-p-q); % steady state probability for regime 0

```

```

ssp1=(1-p)/(2-p-q); % steady state probability for regime 1

```

```

delta0=para(5);

```

```

delta1=para(6);

```

```

sig0=para(7);

```

```

sig1=para(8);

```

```

lam0=para(9);

```

```

lam1=para(10);

```

```

siguu0=para(11);

```

```

sigud0=para(12);

```

```

corr0=para(13);

```

```

varu0=[siguu0^2 siguu0*sigud0*corr0;siguu0*sigud0*corr0 sigud0^2];

```

```

inv_varu0=(varu0)^(-1);

```

```

det_varu0=det(varu0);

```

```

siguu1=para(14);

```

```

sigud1=para(15);

```

```

corr1=para(16);

```

```

varu1=[siguu1^2 siguu1*sigud1*corr1;siguu1*sigud1*corr1 sigud1^2];

```

```

inv_varu1=(varu1)^(-1);

```

```

det_varu1=det(varu1);

```



```

tran_prob=[p 1-q;1-p q]; % transition prob
ssp=[ssp0; ssp1]; % 2 by 1 vector of steady state probability
%% BEGIN ITERATION

lik_value=0;

for jj_iter=3:t-1
    sr_t=sr(jj_iter-1,1); % the short rate which will be used in bondpricing

    % calculate yields for 5 and 10 years bonds using bondpricing
    [yield0,yield1]=bondpricing(sr_t,longest_month_maturity,tran_prob,k0,k1,lam0,lam1,sig0,sig1,delta0,delta1);
    % regime dependent errors (observed short rate - CIR short rate)
    sr_error0=(sr(jj_iter-1,1)-(sr(jj_iter-2,1)+k0*(delta0-sr(jj_iter-2,1))))/sqrt(sr(jj_iter-2,1));
    sr_error1=(sr(jj_iter-1,1)-(sr(jj_iter-2,1)+k1*(delta1-sr(jj_iter-2,1))))/sqrt(sr(jj_iter-2,1));
    % regime dependent likelihood value from short rate equation
    lik_sr0=(sqrt(2*pi)*sig0*sqrt(sr(jj_iter-2,1)))\...
        exp((-sr_error0^2)/(2*sig0^2));
    lik_sr1=(sqrt(2*pi)*sig1*sqrt(sr(jj_iter-2,1)))\...
        exp((-sr_error1^2)/(2*sig1^2));
    % regime dependent errors (observed 5 year yield - short rate driven 5 year yield)
    ly5y_error0=ly(jj_iter-1,5)-yield0(5*12);
    ly5y_error1=ly(jj_iter-1,5)-yield1(5*12);
    % regime dependent errors (observed 10 year yield - short rate driven 10 year yield)
    ly10y_error0=ly(jj_iter-1,10)-yield0(10*12);
    ly10y_error1=ly(jj_iter-1,10)-yield1(10*12);

    ly_error0=[ly5y_error0 ly10y_error0];
    ly_error1=[ly5y_error1 ly10y_error1];
    % regime dependent likelihood value from 5 and 10 year bond equations
    lik_ly0=((2*pi)^(size(ly_error0,2)/2))*(det_varu0^0.5)\...
        exp(-0.5*ly_error0*inv_varu0*(ly_error0));
    lik_ly1=((2*pi)^(size(ly_error1,2)/2))*(det_varu1^0.5)\...
        exp(-0.5*ly_error1*inv_varu1*(ly_error1));
    % total regime dependent likelihood value
    lik_val0=lik_sr0*lik_ly0; % 2 by 2
    lik_val1=lik_sr1*lik_ly1; % 2 by 2

    % the joint probability which is the steady state prob times the
    % conditional transition prob
    vec_tran_prob=reshape(tran_prob,4,1); % vectorize transition prob to 4 by 1
    vec_ssp=reshape([ssp';ssp'],4,1); % vectorize ssp to 4 by 1
    joint_prob=vec_ssp.*vec_tran_prob; % 4 by 1

    % likelihood times joint probabilities
    lik_val_with_prob=[lik_val0;lik_val0;lik_val1;lik_val1].*joint_prob; % 4 by 1
    % likelihood value which includes all cases
    lik_val=sum(lik_val_with_prob);

    if lik_val>0
        lik_value=lik_value+log(lik_val);
        % update the steady state probabilities
        updated_prob=lik_val_with_prob./lik_val;
        ssp=updated_prob(1:2)+updated_prob(3:4);
    else
        lik_value=lik_value-10;
        ssp=[0.5;0.5];
    end
    minus_lik_value=-lik_value;
end

```

```

function [stl, stt, fit_5y, fit_10y, fit_3m] = hfilter(para_in)
% This function filter out the switching probabilities
global sr ly t longest_month_maturity
para = trans_para(para_in);

% assign parameters
k0 = para(1);
k1 = para(2);

p = para(3);
q = para(4);
ssp0 = (1-q)/(2-p-q); % steady state probability for regime 0
ssp1 = (1-p)/(2-p-q); % steady state probability for regime 1

delta0 = para(5);
delta1 = para(6);

sig0 = para(7);
sig1 = para(8);

lam0 = para(9);
lam1 = para(10);

siguu0 = para(11);
sigud0 = para(12);
corr0 = para(13);
varu0 = [siguu0^2 siguu0*sigud0*corr0; siguu0*sigud0*corr0 sigud0^2];
inv_varu0 = (varu0)^(-1);
det_varu0 = det(varu0);

siguu1 = para(14);
sigud1 = para(15);
corr1 = para(16);
varu1 = [siguu1^2 siguu1*sigud1*corr1; siguu1*sigud1*corr1 sigud1^2];
inv_varu1 = (varu1)^(-1);
det_varu1 = det(varu1);

tran_prob = [p 1-q; 1-p q]; % transition prob
ssp = [ssp0; ssp1]; % 2 by 1 vector of steady state probability
%% BEGIN ITERATION
for jj_iter = 3:t-1
    sr_t = sr(jj_iter-1, 1); % the short rate which will be used in bondpricing

    % calculate yields for 5 and 10 years bonds using bondpricing
    [yield0, yield1] = bondpricing(sr_t, longest_month_maturity, tran_prob, k0, k1, lam0, lam1, sig0, sig1, delta0, delta1);
    % regime dependent errors (observed short rate - CIR short rate)
    sr_error0 = (sr(jj_iter-1, 1) - (sr(jj_iter-2, 1) + k0*(delta0 - sr(jj_iter-2, 1))))/sqrt(sr(jj_iter-2, 1));
    sr_error1 = (sr(jj_iter-1, 1) - (sr(jj_iter-2, 1) + k1*(delta1 - sr(jj_iter-2, 1))))/sqrt(sr(jj_iter-2, 1));
    % regime dependent likelihood value from short rate equation
    lik_sr0 = (sqrt(2*pi)*sig0*sqrt(sr(jj_iter-2, 1)))\...
        exp((-sr_error0^2)/(2*sig0^2));
    lik_sr1 = (sqrt(2*pi)*sig1*sqrt(sr(jj_iter-2, 1)))\...
        exp((-sr_error1^2)/(2*sig1^2));
    % regime dependent errors (observed 5 year yield - short rate driven 5 year yield)
    ly5y_error0 = ly(jj_iter-1, 5) - yield0(5*12);

```

```

ly5y_error1=ly(jj_iter-1,5)-yield1(5*12);
% regime dependent errors (observed 10 year yield - short rate driven 10 year yield)
ly10y_error0=ly(jj_iter-1,10)-yield0(10*12);
ly10y_error1=ly(jj_iter-1,10)-yield1(10*12);

ly_error0=[ly5y_error0 ly10y_error0];
ly_error1=[ly5y_error1 ly10y_error1];
% regime dependent likelihood value from 5 and 10 year bond equations
lik_ly0=((2*pi)^(size(ly_error0,2)/2))*(det_varu0^0.5)\...
    exp(-0.5*ly_error0*inv_varu0*(ly_error0'));
lik_ly1=((2*pi)^(size(ly_error1,2)/2))*(det_varu1^0.5)\...
    exp(-0.5*ly_error1*inv_varu1*(ly_error1'));
% total regime dependent likelihood value
lik_val0=lik_sr0*lik_ly0; % 2 by 2
lik_val1=lik_sr1*lik_ly1; % 2 by 2

% the joint probability which is the steady state prob times the
% conditional transition prob
vec_tran_prob=reshape(tran_prob,4,1); % vectorize transition prob to 4 by 1
vec_ssp=reshape([ssp';ssp'],4,1); % vectorize ssp to 4 by 1
joint_prob=vec_ssp.*vec_tran_prob; % 4 by 1
%=====filter=====
stl_temp=joint_prob(1:2)+joint_prob(3:4);
stl(jj_iter-1,:)=stl_temp';
fit_5y_temp=[yield0(5*12) yield1(5*12)]*stl_temp;
fit_10y_temp=[yield0(10*12) yield1(10*12)]*stl_temp;
fit_5y(jj_iter-1,1)=fit_5y_temp;
fit_10y(jj_iter-1,1)=fit_10y_temp;
fit_3m_temp=[yield0(3) yield1(3)]*stl_temp;
%fit_3m_temp=[(sr(jj_iter-1,1)-sr_error0) (sr(jj_iter-1,1)-sr_error1)]*stl_temp;
fit_3m(jj_iter-1,1)=fit_3m_temp;
%=====
% likelihood times joint probabilities
lik_val_with_prob=[lik_val0;lik_val0;lik_val1;lik_val1].*joint_prob; % 4 by 1
% likelihood value which includes all cases
lik_val=sum(lik_val_with_prob);
%if jj_iter>=10
if lik_val>0
    lik_value=lik_value+log(lik_val);
    % update the steady state probabilities
    updated_prob=lik_val_with_prob./lik_val;
    ssp=updated_prob(1:2)+updated_prob(3:4);
else
    lik_value=lik_value-10;
    ssp=[0.5;0.5];
end
%=====filter=====
stt(jj_iter-1,:)=ssp';
%=====
end

```

Chapter V Joint Modelling of the Nominal and Real Term Structure of Interest Rates and the relation between expected inflation rates and stock return risk premium

The expected rate of inflation is an important economic variable to both policy makers and investors. On the one hand, it measures the credibility of central bank's monetary policy aiming at anchoring the long term inflation rate around its target level. On the other hand, the inflation risk premia demanded by investors to take on the risks of unexpected inflation shocks in nominal bonds investment, if non-negligible, would contaminate the interpretation of the break-even inflation rate¹² as a measure of central bank's credibility that are widely used. However, due to the unobserved nature of real interest rates, expected inflation rates and the inflation risk premia, it is difficult, in reality, to make correct interpretation of these quantities.

The first part of this study focuses on the joint modelling of the nominal and real term structure of interest rates using both UK's nominal bonds and inflation index-linked bonds. To decompose the nominal term structure of interest rates into real interest rates, expected inflation rates and inflation risk premia, we estimate an three-factor essentially affine no-arbitrage term structure model as suggested by Duffee (2002). The second part of this study focuses on the relation between expected inflation rates and stock return risk premium. In a smooth transition vector autoregressive specification, we test whether this relation is linear or nonlinear.

This chapter is organized as follows. Section 1 states the motivation of this study and reviews the related literature. Section 2 introduces the inflation index linked bonds with emphasis placed on the UK market. Section 3 derives the essentially affine term structure of the nominal and real interest rates. Section 4 presents the descriptive statistics of the data used in this study. In section 5, we give a preliminary analysis on the UK's nominal and real rates using principal components analysis and simple regressions. Section 6 explains the issues in estimation and section 7 presents the estimation results. In section 8, we test the nonlinearity in the relation between expected inflation rate and stock return risk premium. Finally, section 9 concludes.

¹² The break-even inflation rate is the difference between observed nominal yields extracted from nominal bonds and real yields extracted from inflation index linked bonds with equivalent maturities.

V.1 Motivation and Literature Review

Many early macroeconomic studies on expected inflation rates assume a constant real interest rate (for example, Fama (1975)), whereby the expected rate of inflation can be derived easily by subtracting the real yield from its nominal counterpart. However, the constant real interest rate hypothesis is often rejected, which forces researchers to adopt various assumptions on real interest rate modelling. To list a few, these includes the mean-reverting real interest rate of Hamilton (1985), the unit root behaviour of Rose (1988), fractional integration of Lai (1997) and Markov switching of Garcia and Perron (1996)¹³.

If, however, we are able to infer the real interest rates from observed inflation index linked bonds, the expected rate of inflation can be derived from the called the break-even inflation rate (which is the difference between the nominal interest rate and real interest rate). Although this approach is widely used amongst economists to infer the expected rate of inflation, the break-even inflation rate is contaminated by the premia demanded by investors who wish to take on the inflation risk (the so-called inflation risk premia). In literature, there is little agreement on the size and sign of inflation risk premia. Nevertheless, the statistically significant inflation risk premia begin to attract more and more research attention in this area. While some studies use hypothetical real interest rates derived from observed inflation rates or expected rates of inflation from surveys, others use yields on real bonds together with yields on nominal bonds to derive the implied expected inflation rate and inflation risk premia.

Whether the inflation risk premia should be zero or not, is related to the well known theory of interest rates due to Fisher (1930). If it is true, as Fisher suggests, the inflation risk premia should be zero, there would be no expected benefits for the governments to issue inflation protected securities, because the inflation surprises should cancel out over time, unless the systematic bias in expected inflation exists. Indeed, according to the literature we reviewed (see Table V-12 in Appendix V), inflation risk premia is either positive or negative in the US, the UK and the Euro area. For example, Campbell and Shiller (1996) found the inflation risk premia to be 0.7% to 1% for a nominal bond with 5 years maturity in the UK. The subsequent works by Evans (1998), Remolona, et al. (1998) and Shen (1998) confirmed Campbell and Shiller's but reported a slightly smaller premia. Risa (2001) using inflation index-linked UK gilts spanning from 1983 to 1999 reported a 1.84% premia for a bond with 5 years maturity. A recent study by Joyce, et al. (2010), although does not report the average inflation risk premia, but plotted the dynamics of it over time. They confirm that inflation risk primia is mostly

¹³ For a more recent survey, see Christopher and David (2008).

positive and time-varying. In the US market, Buraschi and Jiltsov (2005) using data on nominal yields, money supply and price level, reported the average inflation risk premia is 0.7%. Grishchenko and Huang (2008), in their recent study, however, reported negative premia before 2004, which they explained may due to the ignorance of modelling liquidity risk premia in their study for TIPS data pre-2004.

It is easy to link the inflation risk premia to the uncertainty or volatility of inflation, and it is intuitive to understand the inflation risk premia should be positive if the inflation rate is high in uncomfortable economic conditions. The modern theory of finance requires a positive correlation between the real pricing kernel and inflation rate. This is because the pricing kernel is usually high when the economy is in perturbed conditions, as risk-averse investors want to smooth their consumption by moving part of their consumption in good state to bad state. The coincidence of high real pricing kernel and high inflation would just make the inflation risk premia positive. However, the correlation between the pricing kernel and the inflation rate (or the consumption path of investors and inflation rate) may vary through time and hence the dynamics of inflation risk premia over time.

To derive the inflation risk premia and expected rate of inflation simultaneously, we follow Evans (1998,2002) Risa (2001) to use the Fisher equation, which relates the one period nominal yield to one period real interest rate and one period expected rate of inflation such that $y_t^{(1)} = y_t^{*(1)} + E_t(\pi_{t+1})$. Extending this relation to a multi-period bond yield, using (III.3) in section III.2 of Chapter III, we have

$$\begin{aligned}
 y_t^{(n)} &= \frac{1}{n} \sum_{i=0}^{n-1} E_t(y_{t+i}^{(1)}) + \Lambda_t^{(n),y} \\
 &= \frac{1}{n} \sum_{i=0}^{n-1} E_t(y_{t+i}^{*(1)} + E_{t+i}(\pi_{t+i+1})) + \Lambda_t^{(n),y} \\
 &= \frac{1}{n} \sum_{i=0}^{n-1} E_t(y_{t+i}^{*(1)}) + \frac{1}{n} \sum_{i=0}^{n-1} E_t(\pi_{t+i+1}) + \Lambda_t^{(n),y} \\
 &= y_t^{*(n)} + \frac{1}{n} \sum_{i=0}^{n-1} E_t(\pi_{t+i+1}) + \Lambda_t^{(n),y} - \Lambda_t^{*(n),y} \\
 &= y_t^{*(n)} + \frac{1}{n} \sum_{i=0}^{n-1} E_t(\pi_{t+i+1}) + \Lambda_t^{(n),\pi},
 \end{aligned} \tag{V.1}$$

where $y_t^{*(n)}$ is the real yield of a n-period bond, $\frac{1}{n} \sum_{i=0}^{n-1} E_t(\pi_{t+i+1})$ is the n-period average of one-period expected inflation rate and $\Lambda_t^{(n),\pi}$ is the inflation risk premia at time t . The second line reveals the Fisher equation that one-period nominal interest rate is equivalent to the sum of

one-period real interest rate and expected rate of inflation one-period ahead. The third line, which uses the law of iterative expectation on future rate of inflation, depicts that nominal interest rates is the sum of the average expected one-period real interest rate, average expected inflation rate over the maturity of a n-period bond and the n-period nominal term premia. The fourth line assumes the expectation hypothesis on real interest rate holds, which says that the n-period real interest rate is the average future one-period real interest rate plus a real term premia. The fifth line derives the n-period inflation risk premia as the difference between the n-period nominal and real term premia.

As we can see, one of the crucial task of deriving the inflation risk premia is the estimation of both nominal and real term premia that are embedded in nominal and real term structure of interest rates. Recall that the nominal term premia is the extra yield that investors demand to compensate for bearing interest rate risk when holding a long term bond. Consider the gross one-period holding return of a n-period bond, which is the payoff of investing one unit of currency on this bond, can be written as $1 + R_{t+1}^{(n)} = \frac{P_{t+1}^{(n-1)}}{P_t^{(n)}}$. Equivalently, using pricing

kernel we can write $1 = E_t \left(M_{t+1} \left(1 + R_{t+1}^{(n)} \right) \right)$. Taking natural logarithm on both sides of the above equation (see Cochrane and Piazzesi (2005) and Piazzesi (2010)), we have

$$\begin{aligned} E_t \left(p_{t+1}^{(n-1)} - p_t^{(n)} \right) &= -E_t \left(m_{t+1} \right) - \text{cov}_t \left(m_{t+1}, p_{t+1}^{(n-1)} \right) - \frac{1}{2} \left(\text{var}_t \left(m_{t+1} \right) + \text{var}_t \left(p_{t+1}^{(n-1)} \right) \right) \\ &= y_t^{(1)} - \text{cov}_t \left(m_{t+1}, p_{t+1}^{(n-1)} \right) - \frac{1}{2} \text{var}_t \left(p_{t+1}^{(n-1)} \right) \end{aligned} \quad (\text{V.2})$$

where the second equality uses fact $y_t^{(1)} = -E_t \left(m_{t+1} \right) - \frac{1}{2} \text{var}_t \left(m_{t+1} \right)$. In relation to (III.1) in section III.2 of Chapter III, equation (V.2) states the expected one-period return of an n-period nominal bond equals the one-period risk-free interest rate ($y_t^{(1)}$) plus a term premia ($-\text{cov}_t \left(m_{t+1}, p_{t+1}^{(n-1)} \right)$) and minus a Jensen's inequality term ($\frac{1}{2} \text{var}_t \left(p_{t+1}^{(n-1)} \right)$). In other words, the expected one-period excess return of this n-period bond depends on the covariance of the log bond price and the log pricing kernel. If the bond price is negatively correlated with the pricing kernel, investors will demand a positive term premia and vice versa for a positive correlation. To understand the intuition behind this, we substitute back for m_{t+1} in terms of consumption (see Cochrane (2005)), to obtain

$$E_t \left(p_{t+1}^{(n-1)} - p_t^{(n)} \right) - y_t^{(1)} = -\text{cov}_t \left(\log \left(\beta \left(\frac{u'(C_{t+1})}{u'(C_t)} \right) \right), p_{t+1}^{(n-1)} \right) - \frac{1}{2} \text{var}_t \left(p_{t+1}^{(n-1)} \right) \quad (\text{V.3})$$

If $p_{t+1}^{(n-1)}$ is positively correlated with consumption C_{t+1} , which means a negative¹⁴ covariance between marginal utility $u'(C_{t+1})$ and $p_{t+1}^{(n-1)}$, investors will demand higher term premia. This is because the investors prefer to smooth their consumption streams over time, a high asset price at the time when they are already feeling wealthy (a high consumption at that time) will only make their consumption stream more volatile. What makes investors demand lower term premia is crucially depending on the negative correlation between $p_{t+1}^{(n-1)}$ and C_{t+1} , which may help them to smooth their consumption path when they are feeling less wealthy (lower C_{t+1}), in difficult times.

The variation in term premia depends on both the amount of risk and the price of the risk, either of which could change over time due to diverse fundamental variables. For example, the expectations of future inflation, fluctuation in economic activity and different stance of monetary policies could all affect investors' perception of the amount of risk in investing long-term bonds. Meanwhile, investors' risk preference (either apt to risk-averse or risk-loving) would also change over time depending on which business cycle phrases the economy is going through.

In addition to these fundamental economic factors, liquidity premia theory and preferred habitat theory argue that the term premia could also be driven by investors' liquidity considerations and preferred investment horizon. Since long-term bonds are less frequently traded compared to short-term bonds, a liquidity premia is demanded by investors for their liquidity consideration. Certain institutions prefer to invest in certain longer maturities of bonds, a change in the preferred habitat to investing in shorter maturity bonds needs to be compensated by a premia. Furthermore, the market-segmentation theory argues that the bond market is actually split into several segments according to the bond maturities as certain institutions (e.g. pension funds) only invest in long-term bonds while others only participate in short-term bonds market. The whole term structure of interest rates, therefore, is determined by several supply and demand curves from different segmented markets.

¹⁴ Marginal utility of consumption declines with an additional unit of consumption.

Apart from those aforementioned, behavioural reasons may also affect the variation of the term premia. For example, the “flight-to-quality” and “flight-to-liquidity” that happened in unusual market conditions, e.g. the 2007-08 financial crisis period, would lead investors to flying to a safer security class such as the U.S. government bonds. As a result, this increasing demand will eventually drive up bond prices but investors will be willing to accept a lower term premia. Even more, if the crisis intensifies, a flight-to-liquidity phenomenon will further drive up bond prices as the investors prepared to accept a negative term premia. Other factors which may also cause a low or even negative term premia, include, for instance, a large demand on bonds from pension funds and the structural imbalances in international trade which lead to large demand in the U.S. bonds from East Asia countries and oil producing countries.

To be specific on the changes in term premia, one should note that the sign of the term premia can go either positive or negative depending on bond investor’s investment horizon relative to bond supplies and whether it is the real interest rates or inflation drives the interest rate risk. For example, it is generally believed that holding a long-term bond would be riskier if investor’s investment horizon is short, and thus a positive term premia will be required. However, if the investor’s investment horizon is much longer (say 20 years), then holding a 20 year bond would be considered much safer than rolling over the short-term bonds up to 20 years. This implies a negative term premia on the longer term bonds but a positive one on the short-term bonds. Now, consider the effect of an inflation rise on real returns of the investor: rolling over the short-term nominal bonds is much safer than holding a nominal long-term bond since the short-term interest rate would rise one-for-one when inflation rises, yet the real return for holding a long-term nominal bond is less promising and therefore a positive term premia would be demanded. The magnitude of the term premia on the long-term bonds depends on the long-term inflation expectation. Back in the 19th century, when gold standard was adopted, there was little long-term inflation risk and the yield curve was actually sloping downward implying zero or negative term premia on the long-term nominal bonds. However, the era after the breakdown of the Bretton-Woods system in 1970s has seen the rising in long-term inflation risk and thus a positive term premia on long-term nominal bonds. Having noticed the different scenarios we have just discussed, in general, there is hardly any theory can speak on the behaviour of the term premia but only the empirical evidence that shows the time variation of the term premia and different sources that drive it.

V.1.1 Empirical behaviour of nominal term premia in no-arbitrage factor models

Rudebusch, Sack and Swanson (2007) reproduced the term premia on the U.S. nominal 10-year Treasury security using five prevailing methods in estimating the term structure of interest rates. The first method utilize a VAR based model which projects the short rate on three standard macroeconomic variables, including lags of unemployment rate, inflation and the three month treasury bill rate. The horizon dependent forecasts of the short rate are made at each time point, and the average of these forecasts can be used to estimate the risk-neutral long term rate at different maturities. To obtain the relevant term premia at different maturities, one simply subtracts the model implied risk neutral long term rate from the corresponding observed yields. An obvious drawback of this approach is that it does not impose any no-arbitrage conditions to the model parameters, which cannot prevent inconsistency between the yield curves at a given point of time. To circumvent this problem, Ang and Piazzesi (2003) and Bernanke, Reinhart and Sack (2004) imposed the no-arbitrage structure on top of the VAR model in estimating the term premia. Alternatively, Rudebusch and Wu (2008) estimated a macro-finance model which combines a canonical affine no-arbitrage term structure model with a standard New Keynesian macroeconomic model; and Kim and Wright (2005a) employed a standard three-latent-factor no-arbitrage affine term structure model with no underpinning relationships linked to observed macroeconomic variables.

One prominent finding from Rudebusch, *et al.* (2007)'s study is that, all five models estimate a declining 10-year nominal term premia over the sample period, from 5% in early 1980s to 1 % in 2006. Although showing resemblance in the estimated trends for term premia, the volatility of these measures and the correlation between different measures show great uncertainty for choosing different assumptions about the future expected short rates. For example, the lowest estimation of the term premia in 2006 comes from Cochrane and Piazzesi (2005), which is negative 2%. The most stable estimation of the term premia, as claimed by the authors comes from Rudebusch and Wu (2008), which is a positive 2%. This large difference, given the low volatility of the U.S. 10-year nominal yields, highlights the difficulties in estimating the time-varying term premia and illustrates how the estimates would depend crucially on modelling assumptions imposed on different models.

V.1.2 Empirical estimates of inflation risk premia

Due to the unobserved nature of real interest rates, most empirical studies focus on nominal term structure of interest rates. Exceptions like Risa (2001), Hordahl and Tristani (2007), Chernov and Mueller (2008), Grishchenko and Huang (2008), Ang, Bekaert and Wei

(2008b), Haubrich, Pennacchi and Ritchken (2008), D' Amico, Kim and Wei (2010), Christensen, Lopez and Rudebusch (2010), García and Werner (2010) and Joyce, Lildholdt and Sorensen (2010), use nominal bond and inflation data, or in addition, data on inflation index-linked bonds and professional inflation forecasts in survey data.

Inflation index-linked bonds are those whose payoffs are linked to inflation rates or price level. Thus the investments on these bonds are cut out from inflation risks. With the exception of U.K., inflation index-linked bonds in other countries have a fairly short history. The U.S. Treasury introduced its inflation index-linked bonds (called Treasury Inflation Protected Security, or TIPS for short) in early 1997 in an attempt to reduce the Treasury's long-term debt servicing costs. Because of the poor liquidity of the TIPS market in its infancy years, several studies (see D' Amico, *et al.* (2010) and Gurkaynak, Sack and Wright (2010)) reported significant liquidity risk premia estimates well over 1% in the first 4 to 5 years since the inception. This liquidity risk premia, however, declined continuously to below 0.5% in recent years with the lowest occurred in 2005 around 0.1%. In studies of the U.K. inflation index-linked gilts, which includes Remolona, Wickens and Gong (1998), Evans (1998, (2003), Risa (2001) and Joyce, *et al.* (2010), the authors do not explicitly model the liquidity risk premia. Nevertheless, as Joyce, *et al.* (2010) suggested, there might be a flight to quality in recent market disruption, which may temporarily lead to a change in liquidity risk premia.

Apart from using inflation index-linked bonds to estimate the inflation risk premia, one may consider the surveys of inflation forecasts by professional economists and consumers to anchor the expected inflation in Fisher's equation. In the U.S., a number of well known surveys have a lengthy historical record, including the Livingston survey, the Michigan survey and the Survey of Professional Forecasters. Ang, Bekaert and Wei (2007) compared different types of forecasting method to the professional surveys on the forecast of one year ahead annual inflation. They find surveys consistently beat time series models, Philips curve models and term structure models in terms of out-of-sample forecasting performance, which suggests that the estimation of inflation risk premia would benefit from using inflation forecasts in surveys.

Below, we searched the literature for recent estimators of the inflation risk premia, focusing either on the US, the UK or the Euro bond markets. Table V-12 in Appendix V lists some but not a comprehensive list of recent studies on inflation risk premia and various estimates on different maturities. Early evidence on the inflation risk premia was predominantly based on the study of the UK data. Campbell and Shiller (1996) was the first to

use a hypothetical indexed gilts with inflation and real interest rates to estimate the inflation risk premia. Using a sample covers from 1985 to 1994, they found the inflation risk premia to be 0.7% to 1% for a nominal bond with 5 years maturity. The subsequent works by Evans (1998), Remolona, *et al.* (1998) and Shen (1998) confirmed Campbell and Shiller's findings using inflation index-linked gilts and survey data but reported a slightly smaller premia. Risa (2001) applied an essentially affine term structure model using nominal and inflation index-linked UK gilts spanning from 1983 to 1999 reported a 1.84% premia for a bond with 5 years maturity. A recent study by Joyce, *et al.* (2010), although does not report the average inflation risk premia, but plotted the dynamics of it over time. They confirm that inflation risk primia is mostly positive and time-varying.

The use of TIPS in term structure models is originated from Jarrow and Yildirim (2003). In their paper, the HJM term structure model was used to price the inflation index-linked bonds. However, due to the limitation of the HJM framework, there is no inflation risk premia reported in their study. Buraschi and Jiltsov (2005) used a structural model relating the inflation risk premia to the monetary supply and real output. By using data on nominal yields, money supply and price level, they reported the average inflation risk premia is 0.7%. Except from Grishchenko and Huang (2008), which reported negative premia before 2004¹⁵, the picture emerged from other studies is that the inflation risk premia is positive. In addition, although the estimates of premia vary substantially over time, they are inversely related to the length of sample period.

To summarise, the variation in estimating the inflation risk premia across the aforementioned studies certainly reflects not only the different methodologies they adopt but also the different data set they use. The US TIPS data has a relatively shorter sample period than the UK's inflation index-linked gilts. Additionally, the lack of liquidity in the infancy period of TIPS market leads to a contaminated real yield before 2004. Ignoring this liquidity problem would overestimate the real yields and underestimate the inflation risk premia. In contrast to the UK inflation index-linked gilts, the later has the advantage over both Euro and the US TIPS data in terms of having longer time series and fewer problems in liquidity. Although Joyce, *et al.* (2010) mentioned the possible structural break happened in 1992 when the UK adopted an inflation targeting policy, the evidence of current financial crisis remind us

¹⁵ This may due to the ignorance of modelling liquidity risk premia in their study for TIPS data pre-2004. As a result, the real yield estimated is upward biased.

that the permanent structural break, unless convincingly accepted, should always be taken with extra care. Otherwise, the unconditional inflation risk premia should be better studied based on a possibly longer time periods. Alternatively, the regime switching framework of Evans (1998, (2003) and Ang, *et al.* (2008b) would better captures this nonlinearity in the unconditional moments of inflation risk premia¹⁶. Even so, the weakness of Evans (1998, (2003)'s model is that the extended Vasicek (1977) type model, although allows the premia to vary over time, is still rather restrictive: the author in his 1998 paper only reported the existence of an inflation risk premia; the measured premia in his 2003 paper is noisy. On the other hand, Ang, *et al.* (2008b) adopted less restricted model based on Duffee (2002)'s essentially affine term structure model. Yet, they didn't utilise any information from the inflation index-linked bond market.

In our study of the nominal and real term structure of interest rate in the UK gilts, the empirical model is similar to Risa (2001) but we allow less restrictive market price of risk, which shall be determined by the data itself. Meanwhile, we use both nominal and inflation index-linked yields, which are driven purely by three latent factors, to facilitate our estimation. More specifically, we let two factors to determine the dynamics of the real rates and the remaining factor be a extra force to channel the nominal term premia. We discuss more details on the term structure model in section V.3. But before that, we briefly talk about the inflation index-linked gilts in section V.2 and data used in this study in section V.4.

V.2 Inflation index-linked gilts

As its name suggests, the payment of inflation index-linked gilt is linked to the retail price index in the UK, which is a proxy for the price level. Depending on the type of gilt/bond linked to the price level, we have inflation index-linked zero coupon bond, which is considered in many empirical studies, and the capital index-linked bond that is prevailing in bond markets.

Assume perfect indexation for now, the index-linked zero coupon bond is the simplest bond structure, but it is useful as a benchmark for index-linked capital bonds' trading in market. Consider a 10-year nominal zero coupon gilt and a maturity-equivalent inflation index-linked gilt, which pay £100 at maturity in nominal term and real term respectively. If the price levels, at the time of purchasing the gilts and that when the gilts mature, are P_0 and P_{10} . Accordingly, the nominal amounts received when both gilts mature will be £100 and $£100 \times (P_{10}/P_0)$.

¹⁶ Haubrich, *et al.* (2008)'s GARCH volatility model would capture the nonlinearity in conditional moments, but is not sufficient to confine the nonlinear unconditional moments.

Although the index-linked zero coupon bond is rarely observed in market, this hypothetical structure can be derived using capital index-linked bonds in practice.

The main difference between the index-linked zero coupon bond and the capital index-linked bond is that the later pays periodic coupons that are also linked to inflation index. Many governments issue such capital index-linked bonds, for example, United States (TIPS), United Kingdom (ILG), Australia (CAIN series), Canada (RRB), France (OATi and OAT  i) and Italy (BTPE  i). As of 2008, according to the research conducted by Barclays Capital¹⁷, the government issued inflation index-linked bonds comprise \$1.5 trillion of the international debt market.

All inflation index-linked gilts issued by the UK Debt Management Office (DMO) are capital index-linked gilts that pay semi-annual coupons. Unlike the index-linked bonds issued by other countries, an eight-month indexation lag is employed in the UK before September 2005. Specifically, this eight-month indexation lag consists of two months for the compilation and publication of the RPI and a further six months to ensure the accrued interest can be calculated at the start of each coupon period. For example, if the RPI rate published at time t_0 (the start of the current coupon period) is the actual RPI rate at time t_{-2} (RPI normally requires one or two months to publish); investors will know the nominal coupon payment at time t_6 only if the RPI rate at time t_6 is replaced by using the rate at t_{-2} and the RPI rate at time t_0 is replaced with the rate for t_{-10} . As a result, the known nominal coupon payment at time t_6 is calculated as $C_{real} \times (RPI_{t_{-2}} / RPI_{t_{-10}})$, where C_{real} is the real coupon payment guaranteed. This design of indexation lag, which allows investors to know the nominal amount of the next coupon payment, enables a visible calculation of the accrued interest when the gilt changes hands. Although it simplifies the calculation of accrued interest, the weakness of this design is that it has a poor protection against unexpected inflation during the eight-month long period. If an inflation index-linked gilt is unable to perfectly protect investors from unexpected inflation shocks, it suggests an implanted inflation risk premia in the real yield¹⁸. As the development of international inflation index-linked bond market urges a unified design of indexation, the first UK inflation index-linked gilt with a three-month indexation lag was issued in September 2005.

¹⁷ See Barclays Capital Research (2008a).

¹⁸ Although the eight-month indexation lag of the UK inflation index-linked gilt suggest an inflation risk premia rooted in the real yields, Evans (1998) and Risa (2001) respectively reported the magnitude is only 1.5 and 0~4 basis points.

Without knowing the nominal amount of the next coupon payment, accrued interest is calculated based on the cumulative movements of RPI rates running from the last coupon payment date.

Another feature of the inflation index-linked gilt in the UK is its taxation on the uplifts on the inflation compensation, coupon and principal. Before 1996, capital gains from gilts were tax-exempt, whereas the coupon payments were considered as income and hence bear income taxes depending on the receivers' tax regimes. The uplift on the inflation compensation in the UK was regarded as a capital gain and therefore was also tax-exempt. After 1996, however, inflation uplift on coupon is no more tax-exempt, yet the uplift on principal is still tax-free to investors beyond just tax-exempt institutions. Although this change of taxation regime may induce market segmentation and then affect the price of those inflation index-linked gilts, the major buyer, as noted by Risa (2001), is pension funds which are tax-exempt institutions. Therefore, the influence of taxation on the prices of inflation index-linked gilts should be negligible. For more comprehensive description of the inflation index-linked bond in the UK market and other major international bond markets, please refer to Deacon, Derry and Mirfendereski (2004).

V.3 Essentially affine term structure of nominal and real interest rates

In this subsection, we will adopt Duffee (2002)'s essentially affine framework to jointly model the term structures of both nominal and real interest rates. This so-called essentially affine term structure model allows nominal and real yields and the prices of risks to be affine functions of the underlying driving factors. Recent term structure modelling strategies, including Ang and Piazzesi (2003), Ang, *et al.* (2008b) and Hordahl and Tristani (2007) fall into this class of models.

V.3.1 State equation

The state equation describes the dynamics of the unobserved state factors. Litterman and Scheinkman (1991) using principal component analysis find that over 98% of the variation in the returns on the US government fixed-income securities can be explained by three factors, which can be labelled as level, slope and curvature. In this study, we follow the convention in term structure models to assume that there are three unobserved state factors drive the nominal and real term structure of interest rates. We specify the unobserved state factors to follow a VAR(1) process as

$$X_{t+1} = \Phi X_t + \Sigma \varepsilon_{t+1}, \quad (\text{V.4})$$

where X_{t+1} consists of 3 state factors, Φ is a 3-by-3 matrix captures the persistence of the factors and their cross interactions, Σ is a 3-by-3 diagonal matrix depicting each factor's volatility, and ε_{t+1} is a 3-by-1 vector consisting of three independent normally distributed innovations, which each is assumed to have a zero mean and unit variance. Specifically, the state equation parameterization is shown in

$$\begin{bmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{bmatrix} = \begin{bmatrix} \phi_{1,1} & 0 & 0 \\ \phi_{2,1} & \phi_{2,2} & 0 \\ \phi_{3,1} & \phi_{3,2} & \phi_{3,3} \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{bmatrix} + \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t+1} \\ \varepsilon_{2,t+1} \\ \varepsilon_{3,t+1} \end{bmatrix} \quad (\text{V.5})$$

For identification purposes, we restrict Φ to be a lower triangular matrix. By doing this, we allow each factor to be influenced by the previous one.

V.3.2 Nominal and real pricing kernel

The nominal pricing kernel (M_{t+1}) is assumed to be related to the short rate (r_t), the innovation terms to the state factors (ε_{t+1}) and the market price of risk (Λ_t) in the following equation

$$M_{t+1} = \exp\left(-r_t - \frac{1}{2} \Lambda_t' \Lambda_t - \Lambda_t' \varepsilon_{t+1}\right), \quad (\text{V.6})$$

among which the short rate and the market price of risk are both affine functions of the state factors as shown in

$$\begin{aligned} r_t &= \alpha + \beta' X_t \\ \Lambda_t &= \lambda_0 + \lambda_1' X_t \end{aligned} \quad (\text{V.7})$$

where α is a scalar, β and λ_0 are 3-by-1 vectors, and λ_1 is a 3-by-3 matrix.

Now, let M_{t+1}^* represents the real pricing kernel. The real pricing kernel is related to the nominal pricing kernel as $M_{t+1} = \frac{M_{t+1}^*}{1 + \pi_{t+1}}$, which is equivalent to say the nominal pricing kernel is driven by the real pricing kernel discounted by inflation. Analogously, we can write the real pricing kernel, using (V.6), as

$$M_{t+1}^* = \exp\left(-r_t + \pi_{t+1} - \frac{1}{2}\Lambda_t'\Lambda_t - \Lambda_t'\varepsilon_{t+1}\right), \quad (\text{V.8})$$

where

$$\pi_{t+1} = \delta_0 + \delta_1'X_t. \quad (\text{V.9})$$

V.3.3 Nominal bond pricing

Assume the nominal bond price is an exponential affine function of the unobserved state factors as shown in (V.10)

$$P_t^{(n+1)} = \exp(A_{n+1} + B_{n+1}'X_t). \quad (\text{V.10})$$

Apply the terminal condition that $P_t^{(0)} = 1$ for a nominal bond with face value equals 1, it is easily to show that $A_0 = B_0 = 0$. When pricing a nominal bond with one period of maturity left, we can write

$$\begin{aligned} P_t^{(1)} &= E_t(M_{t+1} \times 1) \\ &= E_t\left(\exp\left(-r_t - \frac{1}{2}\Lambda_t'\Lambda_t - \Lambda_t'\varepsilon_{t+1}\right)\right) \\ &= E_t\left(\exp\left(-\alpha - \beta'X_t - \frac{1}{2}\Lambda_t'\Lambda_t - \Lambda_t'\varepsilon_{t+1}\right)\right) \\ &= \exp(-\alpha - \beta'X_t) \end{aligned} \quad (\text{V.11})$$

where the last equality utilises the fact that

$$\begin{aligned} E_t(\exp(-\Lambda_t'\varepsilon_{t+1})) &= \exp\left(E_t(-\Lambda_t'\varepsilon_{t+1}) + \frac{1}{2}\text{var}_t(-\Lambda_t'\varepsilon_{t+1})\right) \\ &= \exp\left(0 + \frac{1}{2}\Lambda_t'\Lambda_t\right). \end{aligned}$$

It is straightforward to see that $A_1 = -\alpha$ and $B_1 = -\beta$. For a nominal bond with $n+1$ periods of maturity left, we have

$$\begin{aligned}
P_t^{(n+1)} &= E_t \left(M_{t+1} P_{t+1}^{(n)} \right) \\
&= E_t \left(\exp \left(-r_t - \frac{1}{2} \Lambda'_t \Lambda_t - \Lambda'_t \varepsilon_{t+1} \right) \exp \left(A_n + B'_n X_{t+1} \right) \right) \\
&= E_t \left(\exp \left(-\alpha - \beta' X_t - \frac{1}{2} \Lambda'_t \Lambda_t - \Lambda'_t \varepsilon_{t+1} \right) \exp \left(A_n + B'_n (\Phi X_t + \Sigma \varepsilon_{t+1}) \right) \right) \\
&= \exp \left(-\alpha - \beta' X_t - \frac{1}{2} \Lambda'_t \Lambda_t + A_n + B'_n \Phi X_t + \frac{1}{2} \text{var}_t \left((-\Lambda'_t + B'_n \Sigma) \varepsilon_{t+1} \right) \right) \\
&= \exp \left(-\alpha - \beta' X_t - \frac{1}{2} \Lambda'_t \Lambda_t + A_n + B'_n \Phi X_t + \frac{1}{2} (\Lambda'_t \Lambda_t + B'_n \Sigma \Sigma' B_n - 2 \Lambda'_t B'_n \Sigma) \right) \\
&= \exp \left(-\alpha - \beta' X_t + A_n + B'_n \Phi X_t + \frac{1}{2} B'_n \Sigma \Sigma' B_n - \Lambda'_t B'_n \Sigma \right) \\
&= \exp \left(-\alpha - \beta' X_t + A_n + B'_n \Phi X_t + \frac{1}{2} B'_n \Sigma \Sigma' B_n - B'_n \Sigma (\lambda_0 + \lambda_1 X_t) \right) \\
&= \exp \left(\left(-\alpha + A_n + \frac{1}{2} B'_n \Sigma \Sigma' B_n - B'_n \Sigma \lambda_0 \right) + (-\beta' + B'_n (\Phi - \Sigma \lambda_1)) X_t \right). \tag{V.12}
\end{aligned}$$

Therefore, the coefficients A_{n+1} and B_{n+1} are obtained from matching the coefficients in (V.12) with those in equation (V.10), which yields

$$\begin{aligned}
A_{n+1} &= A_n - \alpha + \frac{1}{2} B'_n \Sigma \Sigma' B_n - B'_n \Sigma \lambda_0 \\
B'_{n+1} &= -\beta' + B'_n (\Phi - \Sigma \lambda_1). \tag{V.13}
\end{aligned}$$

Subsequently, the nominal yield with maturity of $n+1$ years can be written as an affine function of the state factors as

$$y_t^{(n+1)} = -\frac{1}{n+1} \log \left(P_t^{(n+1)} \right) = a_{n+1} + b'_{n+1} X_t, \tag{V.14}$$

where $a_{n+1} = -\frac{1}{n+1} A_{n+1}$ and $b'_{n+1} = -\frac{1}{n+1} B'_{n+1}$.

V.3.4 Real bond pricing

Let $P_t^{*(n+1)}$ denote the price of a real bond with maturity of $n+1$, we assume it is driven by the same state factors as for the nominal bond, which yields

$$P_t^{*(n+1)} = \exp \left(A_n^* + B_n^{*'} X_t \right). \tag{V.15}$$

Using (V.8) and (V.9), we can derive the price of a real bond as

$$\begin{aligned}
P_t^{*(n+1)} &= E_t \left(M_{t+1}^* P_{t+1}^{*(n)} \right) \\
&= E_t \left(\exp \left(-r_t + \pi_{t+1} - \frac{1}{2} \Lambda_t' \Lambda_t - \Lambda_t' \varepsilon_{t+1} \right) \exp \left(A_n^* + B_n'^* X_{t+1} \right) \right) \\
&= E_t \left(\exp \left(-\alpha - \beta' X_t + \delta_0 + \delta_1' X_t - \frac{1}{2} \Lambda_t' \Lambda_t - \Lambda_t' \varepsilon_{t+1} \right) \exp \left(A_n^* + B_n'^* (\Phi X_t + \Sigma \varepsilon_{t+1}) \right) \right) \\
&= \exp \left((-\alpha + \delta_0) - (\beta' - \delta_1') X_t - \frac{1}{2} \Lambda_t' \Lambda_t + A_n^* + B_n'^* \Phi X_t + \frac{1}{2} \text{var}_t \left((-\Lambda_t' + B_n'^* \Sigma) \varepsilon_{t+1} \right) \right) \\
&= \exp \left((-\alpha + \delta_0) - (\beta' - \delta_1') X_t - \frac{1}{2} \Lambda_t' \Lambda_t + A_n^* + B_n'^* \Phi X_t + \frac{1}{2} \left(\Lambda_t' \Lambda_t + B_n'^* \Sigma \Sigma' B_n^* - 2 \Lambda_t' B_n'^* \Sigma \right) \right) \\
&= \exp \left((-\alpha + \delta_0) - (\beta' - \delta_1') X_t + A_n^* + B_n'^* \Phi X_t + \frac{1}{2} B_n'^* \Sigma \Sigma' B_n^* - \Lambda_t' B_n'^* \Sigma \right) \\
&= \exp \left((-\alpha + \delta_0) - (\beta' - \delta_1') X_t + A_n^* + B_n'^* \Phi X_t + \frac{1}{2} B_n'^* \Sigma \Sigma' B_n^* - B_n'^* \Sigma (\lambda_0 + \lambda_1 X_t) \right) \\
&= \exp \left(\left((-\alpha + \delta_0) + A_n^* + \frac{1}{2} B_n'^* \Sigma \Sigma' B_n^* - B_n'^* \Sigma \lambda_0 \right) + \left(-(\beta' - \delta_1') + B_n'^* (\Phi - \Sigma \lambda_1) \right) X_t \right).
\end{aligned}
\tag{V.16}$$

Analogously, by matching the coefficients in (V.16) with those in (V.15), it yields

$$\begin{aligned}
A_{n+1}^* &= (-\alpha + \delta_0) + A_n^* + \frac{1}{2} B_n'^* \Sigma \Sigma' B_n^* - B_n'^* \Sigma \lambda_0 \\
B_{n+1}'^* &= -(\beta' - \delta_1') + B_n'^* (\Phi - \Sigma \lambda_1).
\end{aligned}
\tag{V.17}$$

The real yield, like the nominal yield, is also an affine function of the state factors and can be written as

$$y_t^{*(n+1)} = -\frac{1}{n+1} \log \left(P_t^{*(n+1)} \right) = a_{n+1}^* + b_{n+1}'^* X_t
\tag{V.18}$$

where $a_{n+1}^* = -\frac{1}{n+1} A_n^*$ and $b_{n+1}'^* = -\frac{1}{n+1} B_n'^*$.

V.3.5 Nominal and real term premia

In affine term structure models, the term premia is primarily driven by the market price of risk. To see the intuition behind this, we notice that from (V.3), we have

$$\begin{aligned}
& \text{cov}_t \left(\log \left(\beta \left(\frac{u'(C_{t+1})}{u'(C_t)} \right) \right), p_{t+1}^{(n-1)} \right) = \text{cov}_t \left(m_{t+1}, p_{t+1}^{(n-1)} \right) \\
& = \text{cov}_t \left(\left(-r_t - \frac{1}{2} \Lambda_t' \Lambda_t - \Lambda_t' \varepsilon_{t+1} \right), (A_{n-1} + B_{n-1} (\Phi X_t + \Sigma \varepsilon_{t+1})) \right) \quad (\text{V.19}) \\
& = -\Lambda_t' B_{n-1} \Sigma \text{cov}_t (\varepsilon_{t+1}, \varepsilon_{t+1}) \\
& = -(\lambda_0 + \lambda_1 X_t)' B_{n-1} \Sigma,
\end{aligned}$$

where m_{t+1} and $p_{t+1}^{(n-1)}$ are natural logarithm of the pricing kernel and bond price respectively. This instantaneous risk premia of investing in a bond with n-period maturity comes from the time-varying market price of risk Λ_t , the maturity of this bond (as represented by coefficient B_{n-1}) and the size of risks is measured by the volatility of underlying unobserved state factors Σ . It is then clearly shown that the instantaneous risk premia increases if either the market price of risk is higher (due to the volatile underlying state factors) or the maturity of the bond is longer. In the case of risk neutrality, however, the risk premia will be zero as the risk neutral investors will not price the risk ($\Lambda_t = 0$).

The nominal term premia, according to the definition in Chapter III, can be constructed as

$$\begin{aligned}
& y_t^{(n)} - \frac{1}{n} \sum_{i=0}^{n-1} E_t (y_{t+i}^{(1)}) \\
& = (a_n + b_n' X_t) - \frac{1}{n} \sum_{i=0}^{n-1} (\alpha + \beta' E_t (X_{t+i})) \quad (\text{V.20}) \\
& = a_n + b_n' X_t - \alpha - \frac{1}{n} \beta' \sum_{i=0}^{n-1} \Phi^i X_t.
\end{aligned}$$

The real term premia, following the same logic, can be derived as

$$\begin{aligned}
& y_t^{*(n)} - \frac{1}{n} \sum_{i=0}^{n-1} E_t (y_{t+i}^{*(1)}) \\
& = (a_n^* + b_n^{*'} X_t) - \frac{1}{n} \sum_{i=0}^{n-1} \left(\alpha - \delta_0 + (\beta - \delta_1)' E_t (X_{t+i}) \right) \quad (\text{V.21}) \\
& = a_n^* + b_n^{*'} X_t - (\alpha - \delta_0) - \frac{1}{n} (\beta - \delta_1)' \sum_{i=0}^{n-1} \Phi^i X_t.
\end{aligned}$$

V.3.6 Inflation risk premia and expected inflation

Having defined both nominal and real term premia, we are now ready to construct the inflation risk premia and the expected inflation. As equation (V.1) shows, the inflation risk premia is simply the difference between the nominal and the real term premia. Therefore, using (V.20) and (V.21), we can construct the inflation risk premia as

$$\begin{aligned}\Lambda_t^{(n),\pi} &= \left(a_n + b'_n X_t - \alpha - \frac{1}{n} \beta' \sum_{i=0}^{n-1} \Phi^i X_t \right) - \left(a_n^* + b_n'^* X_t - (\alpha - \delta_0) - \frac{1}{n} (\beta - \delta_1)' \sum_{i=0}^{n-1} \Phi^i X_t \right) \quad (V.22) \\ &= (a_n - a_n^*) + (b'_n - b_n'^*) X_t - \delta_0 - \delta_1' \frac{1}{n} \sum_{i=0}^{n-1} \Phi^i X_t.\end{aligned}$$

Subsequently, the expected inflation rate, according to (V.1), can be derived as

$$\begin{aligned}\frac{1}{n} E_t \sum_{i=1}^n \pi_{t+i} &= y_t^{(n)} - y_t^{*(n)} - \Lambda_t^{(n),\pi} \\ &= (a_n - a_n^*) + (b'_n - b_n'^*) X_t - (a_n - a_n^*) - (b'_n - b_n'^*) X_t + \delta_0 + \delta_1' \frac{1}{n} \sum_{i=0}^{n-1} \Phi^i X_t \quad (V.23) \\ &= \delta_0 + \delta_1' \frac{1}{n} \sum_{i=0}^{n-1} \Phi^i X_t.\end{aligned}$$

V.4 Data

The nominal and inflation index-linked yields used in this study are obtained from the Bank of England¹⁹. We use the end-of-month zero-coupon bond nominal and real yields produced by the Bank of England. The method they used in calculating the nominal yield curve is essentially a cubic-spline method which is incorporated with a penalty smooth function that allows the yields increase smoothly with maturity. The nominal yield curve is constructed using general collateral repo rates and the coupon-paying nominal government gilts. The real yield curve is based on capital index-linked gilts and the nominal counterparts while implicitly assuming that there is no indexation lag risk premia on the inflation index-linked gilts. The method they used to construct the real yield curve is similar to Evans (1998)²⁰.

¹⁹ The Bank of England publishes nominal and real yield curve estimates, which can be obtained from <http://www.bankofengland.co.uk/statistics/yieldcurve>

²⁰ For more information on the method that the Bank of England used to construct the real and nominal yield curves, please refer to Anderson and Sleath (2001).

The nominal yield data that the Bank of England supplied starts from January 1970, whereas the real yield data begins from January 1985. The short end of the nominal yield curve is missing until the March of 1997. In order to incorporate longer sample period in our study, especially the short end of the yield curve, we filled the missing observations on 3-month interest rate from January 1985 to February 1985 with those obtained from Datastream. As a result, we include nominal yield data ranges from January 1985 to September 2010 for maturities 3-month, and 1-year to 10-year.

The real yield data considered in this study has maturities ranges from 4-year to 10-year starting from January 1985 to September 2010. The shortest maturity on the real yield data that Bank of England offers is 25-month. However, the many missing observations in this maturity series prevent us from using it. The same problem also exists in maturities 26-month to 47-month. Although we have discarded the information contained in the short-end of the real yield curve; it is sufficient for us to use maturities 4-year to 10-year in this study²¹, since the 5-year and 10-year expected inflation and the corresponding inflation risk premia are the most relevant quantities we want to abstract from the nominal and real yield curve.

²¹ Although the 4-year real yield data is the shortest and most complete maturity series, there is still one data point missing on December 1996. To solve this missing data problem, we interpolated this data point linearly by using the neighbouring points.

Figure V-1: UK nominal zero-coupon yields for various maturities

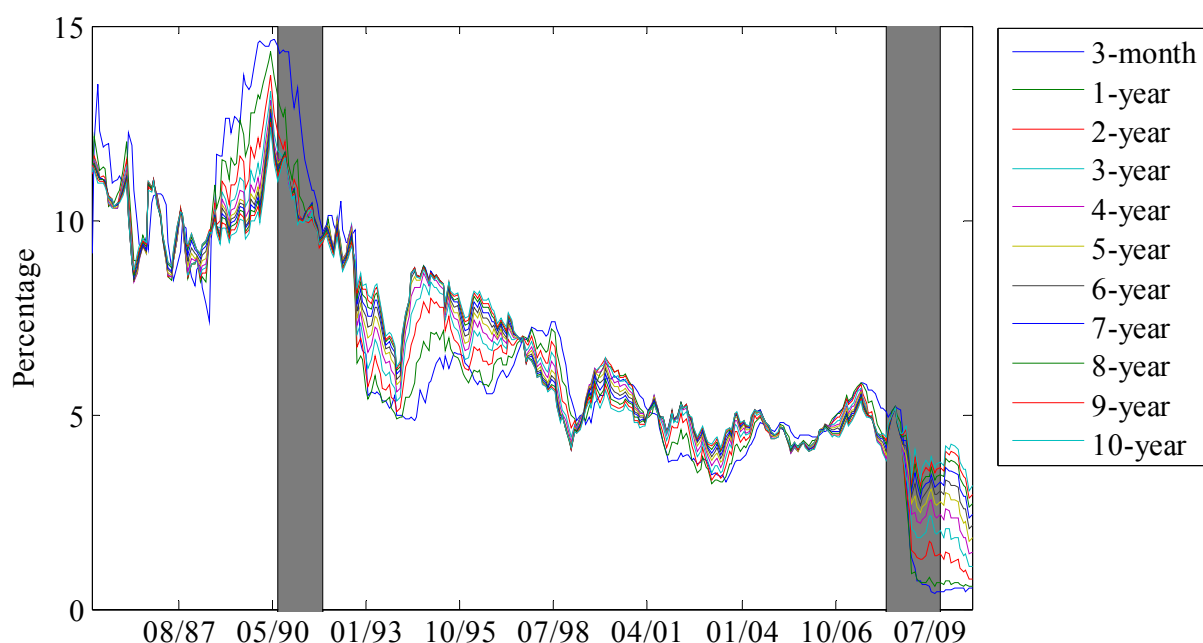
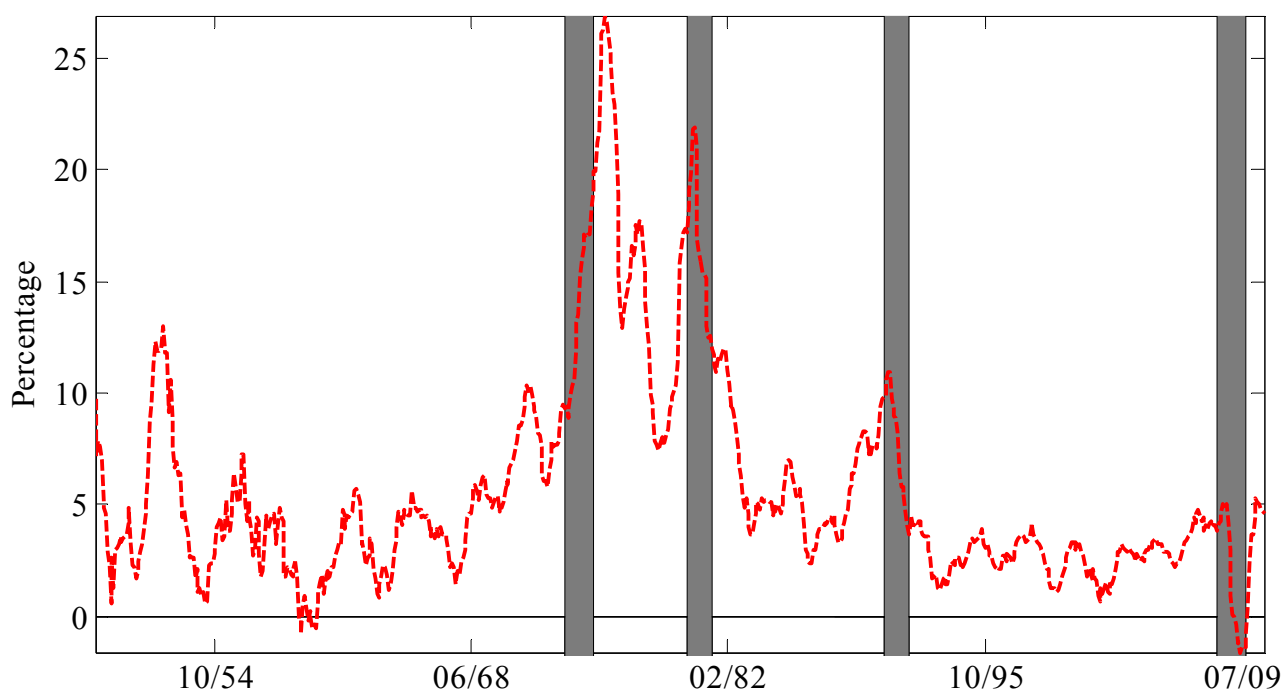


Figure V-1 plots the UK's nominal zero-coupon yields for different maturities from January 1985 to September 2010. The shaded areas represent the UK's recent recessions occurred in July 1990-September 1991 and April 2008-September 2009²². The first prominent feature from inspecting the dynamics of the nominal yields is the declining trend during this time period. At the time where our sample period begins, the UK economy was recovering from the second oil crisis. The high inflation caused by surges of oil prices during the Iranian revolution in 1979 and Iraqi's invasion of Iran in 1980 was brought down from over 25% to around 5% (see the plot of RPI in Figure V-2 Figure V-2 , where the inflation index-linked gilts are tied to RPI). The declining interest rate, however, entered into a rising channel around the 1987's stock market collapse, which caused a inverted yield curve during late 80s and early 90s. The quick recovery from the 1987 crisis soon pushed the annual inflation rate to a level of 10% around 1990, where the UK joined the European Exchange Rate Mechanism (ERM). The 1990s' recession (shown as the third shaded bar in Figure V-2) and high inflation caused by the spikes of oil prices, however, forced the UK's government to exit the ERM programme after the pound sterling came under major pressure from currency speculation two years later. The inflation rate, soon after the exit of the ERM, was officially targeted to a range between 1%

²² The dates of UK's previous recessions can be found in "Inflation Report", Bank of England (2009 Feb). The recent recession from the second quarter of 2008 to the third quarter of 2009 is reported by BBC at <http://news.bbc.co.uk/1/hi/8479639.stm>.

and 4%. This target was further reduced to 2.5% in 1997 when the independence of the Bank of England was granted. The subsequent replacement of retail price index by the consumer price index as the treasury's inflation index in 2003 changed the target to a level of 2%²³. The interest rate after 1995 showed a continuous declining trend until the late 2008 where the recent financial crisis brought a near zero short rate. Notably, the annual inflation rate in 2009 even showed a deflation between March and October in the recent recession.

Figure V-2: UK RPI(Percentage change over 12 months - all items)²⁴



The time series of the UK's real yields for maturities 4-year to 10-year is plotted in Figure V-3. The annual real interest rate has fallen continuously over the sample period with sharp drops around the 1992 ERM event, the 1998 Russia debt crisis and the recent financial crisis. The sharp rise between November and December of 2008, however, is worth noting. According to Bank of England's Inflation Report in February of 2009, the CPI inflation fell to 3.1% in December of 2008, from 5.2% in September, which created a largest three-month fall since 1992. This substantial fall, partially due to the energy price fall in November and the change of VAT regime in December, lead to an expectation of continuous material fall in the

²³ See "Key Monetary Policy Dates Since 1990" at

<http://web.archive.org/web/20070629143630/http://www.bankofengland.co.uk/monetarypolicy/history.htm>

²⁴ The data is obtained from Office for National Statistics at <http://www.statistics.gov.uk/cci/nugget.asp?id=19>

CPI inflation in the coming months. The short-lived deflation between March and October of 2009, as shown in Figure V-2, clearly captured this outlook. Meanwhile, the Monetary Policy Committee's decision of cutting bank rate by 3.5% (since the beginning of November) to 1% would significantly push the real rates higher given the prospective deflation in the near term.

Figure V-3: UK zero-coupon real yields for various maturities

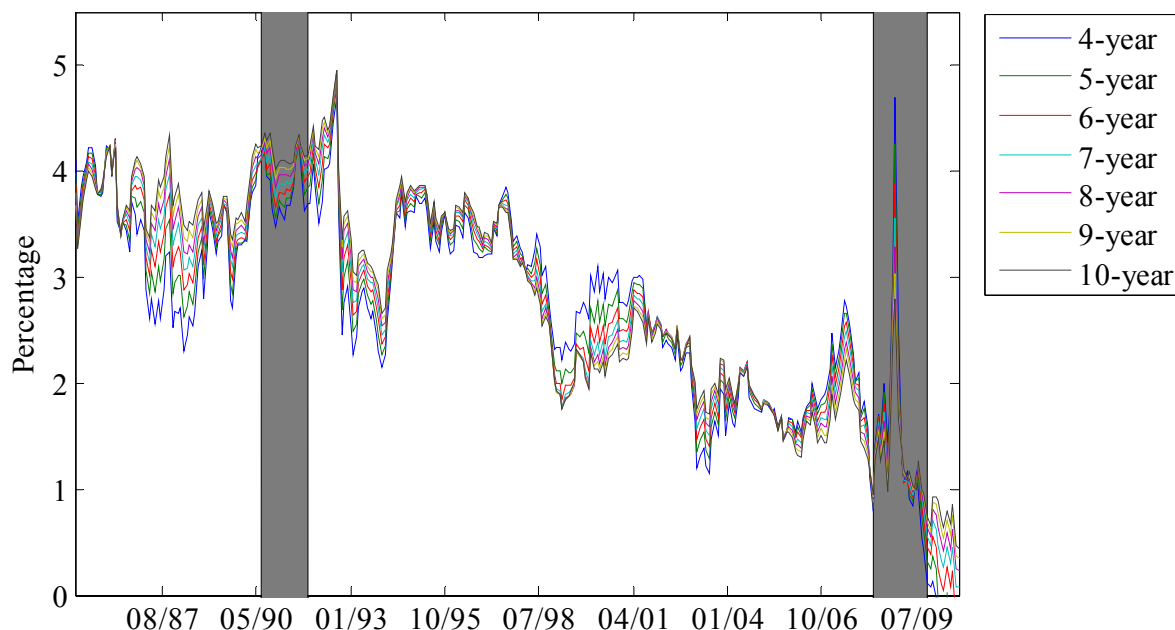


Table V-1 and Table V-2 display the descriptive statistics of the nominal and real yields studied in this paper. One noteworthy feature from the tables is that there are high cross-correlations of yields among the nominal and real curves. This high cross-correlation suggests that a small number of common factors drive the co-movement of the nominal yields across different maturities. Since the correlation coefficients are not exactly unity, this suggests there are non-parallel shifts which cannot be captured by simple linear one-factor models²⁵.

²⁵ We confirm this point with the principle component analysis later.

Table V-1: Descriptive statistics of nominal zero-coupon bond yields

	$y^{(3m)}$	$y^{(1y)}$	$y^{(2y)}$	$y^{(3y)}$	$y^{(4y)}$	$y^{(5y)}$	$y^{(6y)}$	$y^{(7y)}$	$y^{(8y)}$	$y^{(9y)}$	$y^{(10y)}$
Mean	6.717	6.568	6.618	6.689	6.750	6.798	6.837	6.865	6.883	6.891	6.889
Median	5.763	5.850	6.180	6.238	6.302	6.254	6.171	6.133	6.229	6.200	6.249
Maximum	14.67	14.31	13.72	13.31	13.07	12.93	12.82	12.73	12.62	12.50	12.36
Minimum	0.403	0.579	0.771	1.080	1.422	1.757	2.070	2.353	2.606	2.828	3.022
Std. Dev.	3.463	3.138	2.887	2.740	2.652	2.597	2.558	2.528	2.501	2.473	2.443
Skewness	0.532	0.332	0.238	0.225	0.241	0.264	0.285	0.300	0.309	0.312	0.310
Kurtosis	2.763	2.654	2.409	2.205	2.054	1.941	1.852	1.779	1.719	1.667	1.621
Normality Test											
Jarque-Bera	15.26	7.22	7.42	10.76	14.52	18.04	21.15	23.81	26.05	27.90	29.44
Probability	0.000	0.027	0.024	0.005	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Correlation Matrix											
$y^{(3m)}$	1	0.983	0.966	0.950	0.937	0.925	0.915	0.905	0.896	0.887	0.878
$y^{(1y)}$		1	0.993	0.982	0.970	0.959	0.948	0.938	0.929	0.919	0.910
$y^{(2y)}$			1	0.997	0.990	0.983	0.975	0.967	0.958	0.950	0.942
$y^{(3y)}$				1	0.998	0.994	0.989	0.983	0.976	0.970	0.963
$y^{(4y)}$					1	0.999	0.996	0.992	0.987	0.982	0.976
$y^{(5y)}$						1	0.999	0.997	0.994	0.990	0.985
$y^{(6y)}$							1	0.999	0.997	0.995	0.991
$y^{(7y)}$								1	0.999	0.998	0.996
$y^{(8y)}$									1	1.000	0.998
$y^{(9y)}$										1	1.000
$y^{(10y)}$											1
$y^{*(4y)}$	0.743	0.781	0.811	0.819	0.817	0.809	0.798	0.787	0.776	0.766	0.758
$y^{*(5y)}$	0.768	0.806	0.841	0.855	0.856	0.852	0.846	0.838	0.830	0.823	0.816
$y^{*(6y)}$	0.781	0.820	0.858	0.875	0.880	0.879	0.875	0.870	0.864	0.859	0.855
$y^{*(7y)}$	0.788	0.826	0.867	0.886	0.894	0.896	0.894	0.891	0.887	0.883	0.880
$y^{*(8y)}$	0.791	0.830	0.871	0.892	0.902	0.906	0.906	0.904	0.901	0.899	0.897
$y^{*(9y)}$	0.793	0.831	0.873	0.896	0.907	0.912	0.913	0.912	0.911	0.909	0.908
$y^{*(10y)}$	0.793	0.830	0.873	0.897	0.909	0.915	0.917	0.918	0.917	0.916	0.915

Table V-2: Descriptive statistics of real zero-coupon bond yields

	$y^{*(4y)}$	$y^{*(5y)}$	$y^{*(6y)}$	$y^{*(7y)}$	$y^{*(8y)}$	$y^{*(9y)}$	$y^{*(10y)}$
Mean	2.678	2.712	2.739	2.761	2.779	2.793	2.804
Median	2.779	2.853	2.882	2.941	2.945	2.962	2.970
Maximum	4.703	4.801	4.869	4.913	4.937	4.947	4.945
Minimum	-0.749	-0.388	-0.121	0.083	0.228	0.341	0.430
Std. Dev.	1.046	1.018	1.013	1.020	1.031	1.044	1.059
Skewness	-0.895	-0.707	-0.551	-0.434	-0.350	-0.289	-0.246
Kurtosis	3.913	3.236	2.723	2.372	2.138	1.980	1.872
Normality Test							
Jarque-Bera	51.982	26.482	16.629	14.788	15.863	17.691	19.504
Probability	0.000	0.000	0.000	0.001	0.000	0.000	0.000
Correlation Matrix							
$y^{*(4y)}$	1	0.991	0.972	0.951	0.930	0.912	0.895
$y^{*(5y)}$		1	0.995	0.983	0.970	0.957	0.944
$y^{*(6y)}$			1	0.997	0.990	0.981	0.972
$y^{*(7y)}$				1	0.998	0.993	0.988
$y^{*(8y)}$					1	0.999	0.996
$y^{*(9y)}$						1	0.999
$y^{*(10y)}$							1

The average nominal yield curve with a 95% confidence interval for the sample period we studied is plotted in Figure V-4. In general, the yield curve is sloping upwards from 1-year maturity to 10-year maturity. Yet, yield with a maturity of 3-month is modestly higher. A 10 basis points difference between the 3-month and 1-year maturities creates a kinked shape at the short end. This may due to the volatile movements of the short rate at the beginning of the sample period, caused by a series of high inflation and recessions periods in the late 80s and early 90s. Figure V-6 shows the average real yield curve for the sample period between January 1985 and September 2010. An upward sloping yield curve gives an average yield of 2.7% for a 5-year real bond and 2.8% for a 10-year real bond.

The term structure of the volatility (or volatility curve) of the nominal yields is shown in Figure V-5. Unlike the "snake" shape of the volatility curve found in the US nominal yields, the UK's curve is downward sloping with maturities for the sample period studied in this paper. The similar shape of the volatility curve in the real yields is plotted in Figure V-7. These

smooth declining volatilities along the maturities would suggest that the use of multifactor affine models with independent factors will suffice to produce the observed volatility curve²⁶.

Figure V-4: UK average nominal yield curve with +/- 2 S.E. bounds

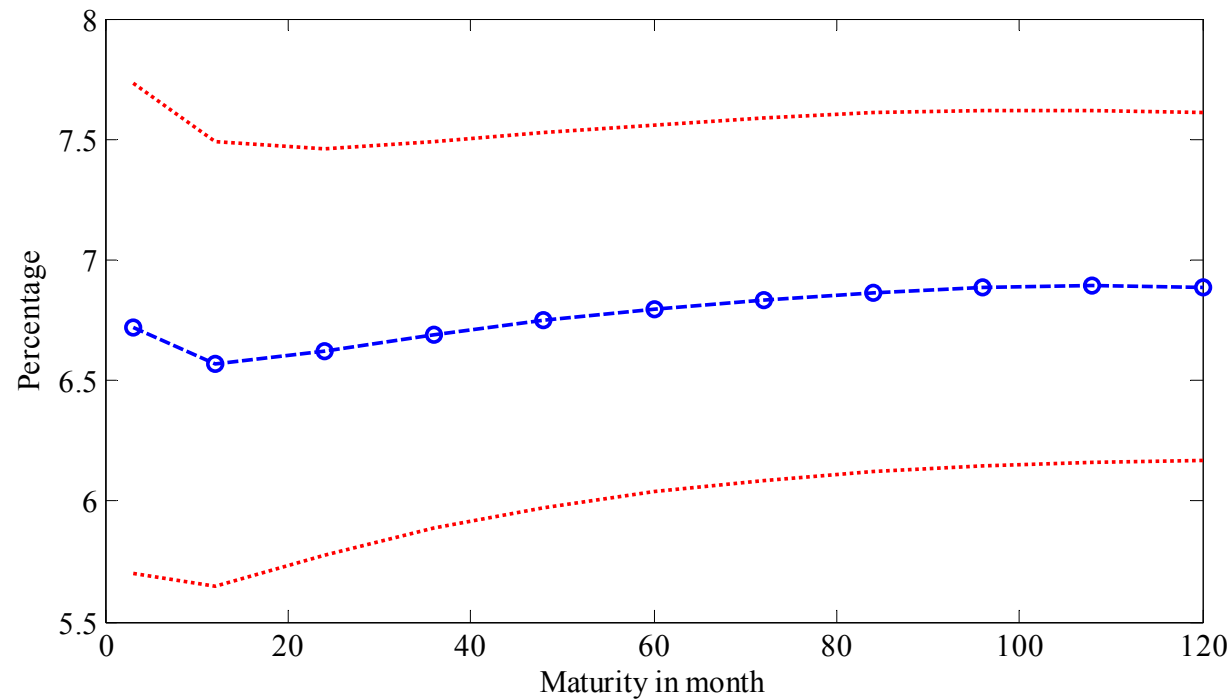
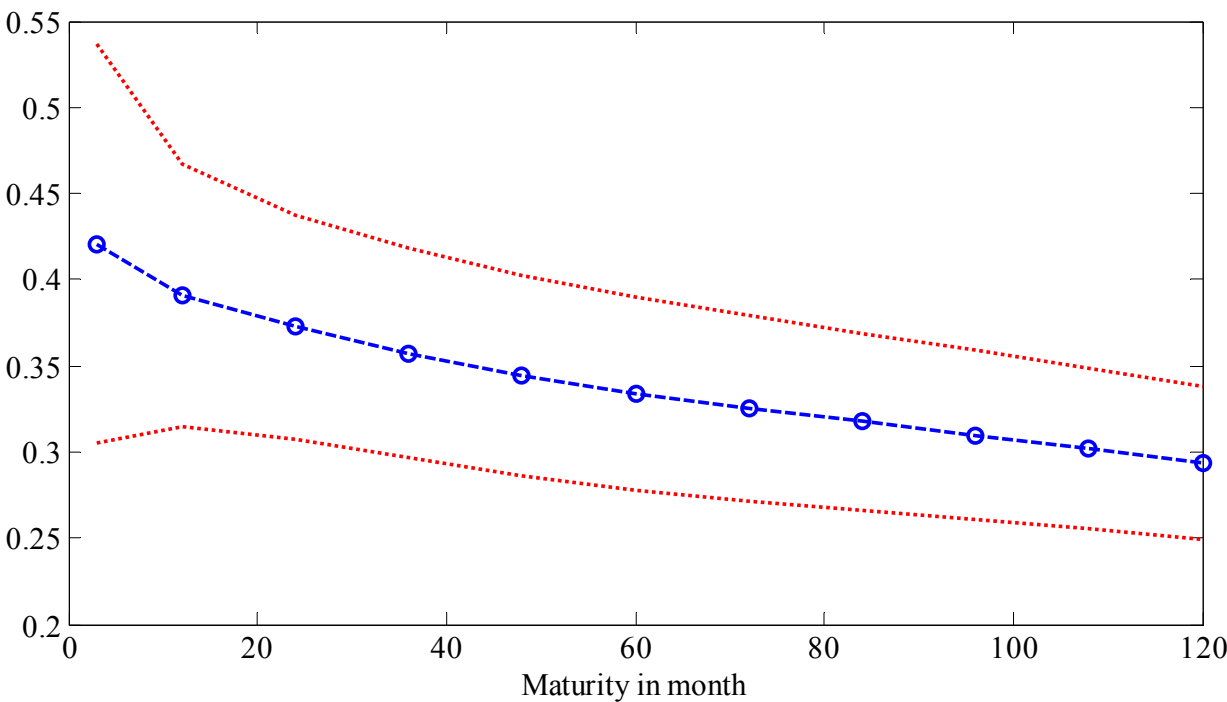


Figure V-5: UK nominal yield volatility curve with +/- 2 S.E. bounds



²⁶ As documented in Piazzesi (2010), the "snake" shaped volatility curve found in the US nominal yields requires negatively correlated factors to produce.

Figure V-6: UK average real yield curve with ± 2 S.E. bounds

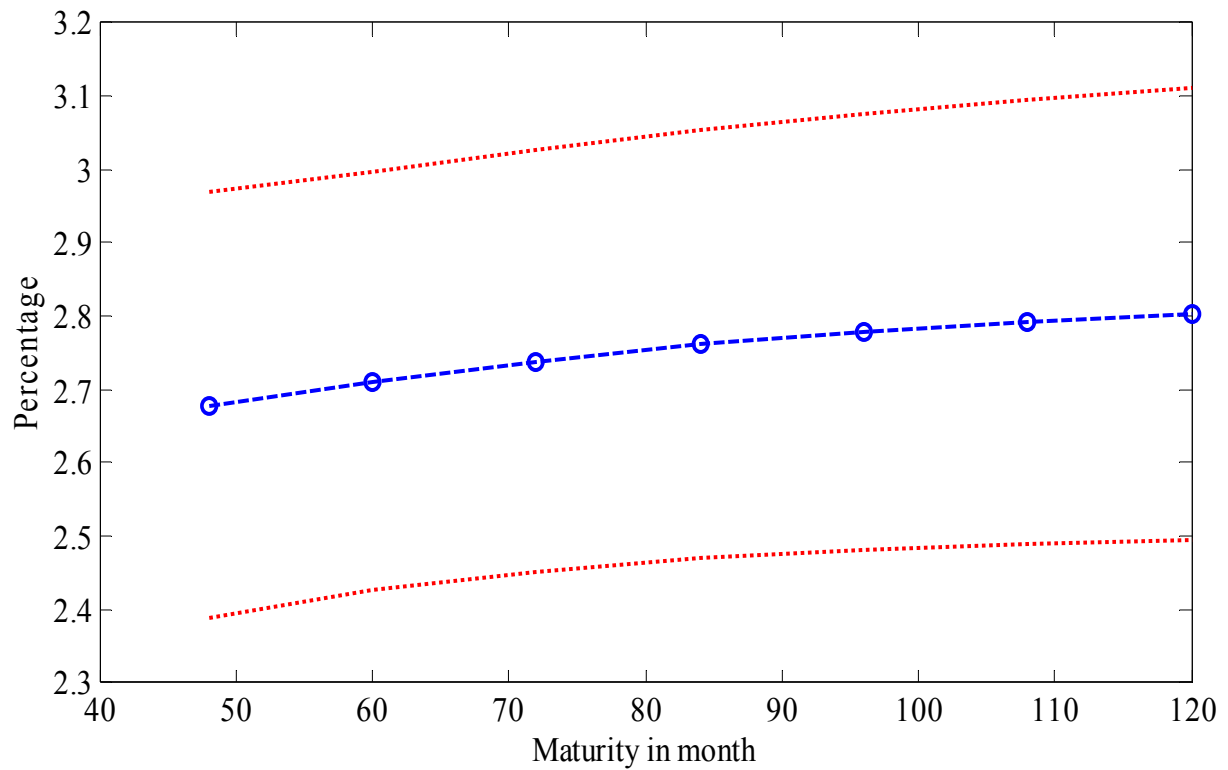
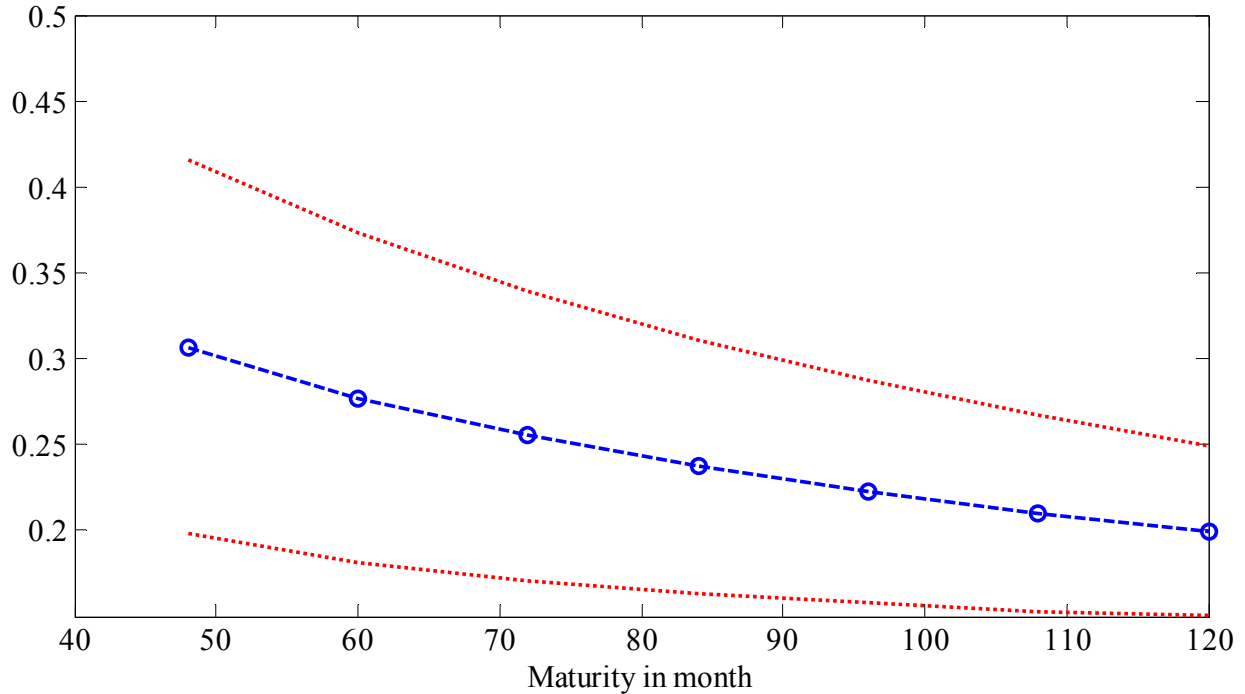


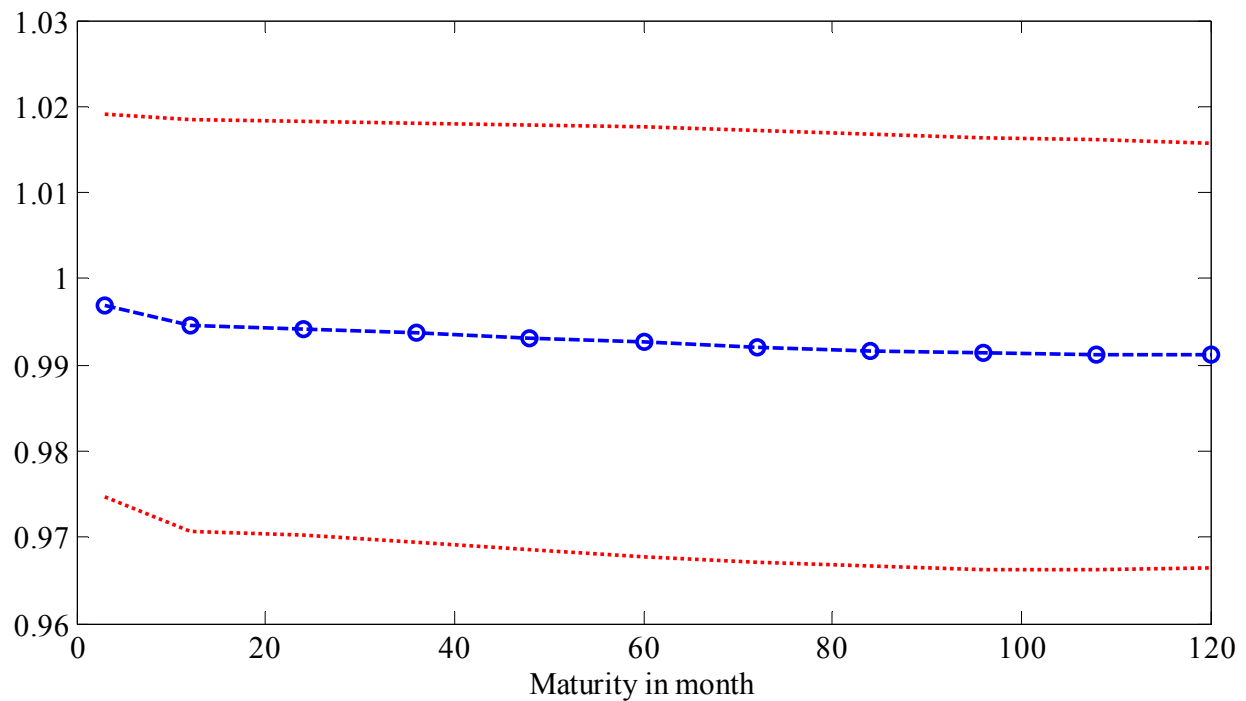
Figure V-7: UK real yield volatility curve with ± 2 S.E. bounds



Previous studies on the US nominal yield curve found yields are highly persistent. Figure V-8 and Figure V-9 show the autocorrelation curves of the UK nominal and real yields

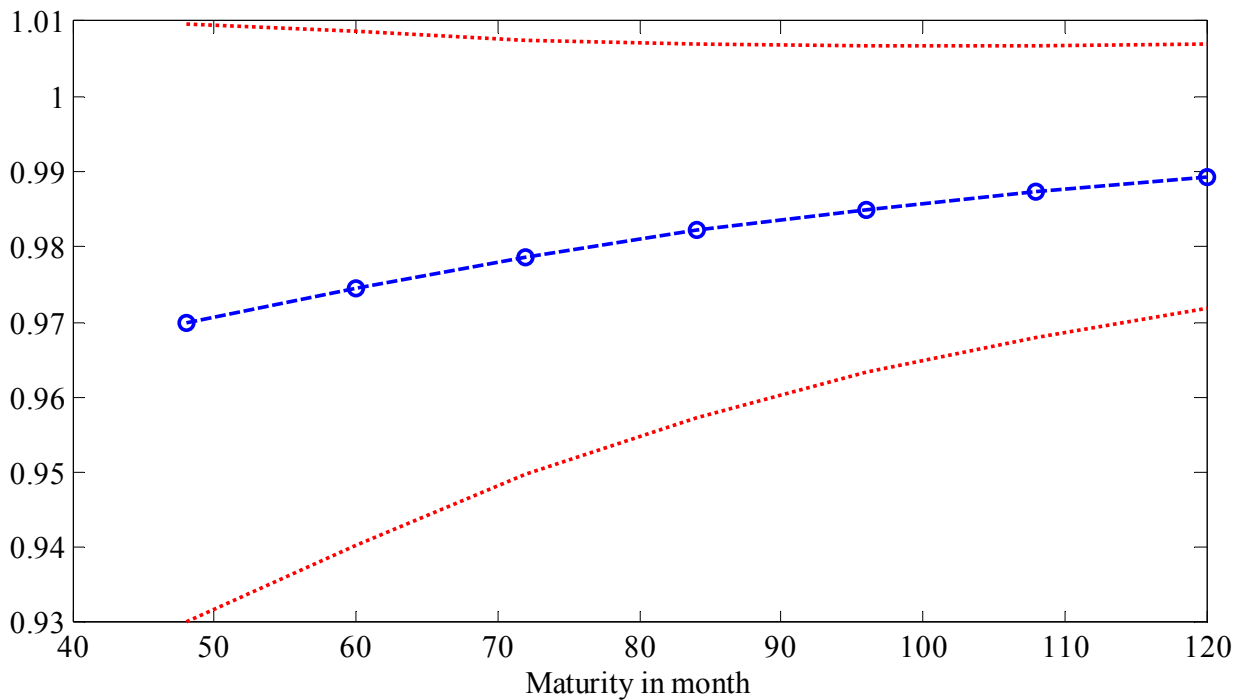
together with standard errors around the estimates²⁷. It is interesting to see the monthly autocorrelation coefficients for the nominal yields are slightly decreasing with maturities, which implicitly suggests the shorter maturity yields have a slightly longer half-life than those at the long ends. This high persistence in nominal and real yields, self-explains the large standard errors around their mean estimates in Figure V-4 and Figure V-6.

Figure V-8: UK nominal yield autocorrelation curve with +/- 2 S.E. bounds



²⁷ These autocorrelation estimates are not corrected for small sample bias, standard errors are calculated with 6 Newey West lags.

Figure V-9: UK real yield autocorrelation curve with ± 2 S.E. bounds



V.5 Principal component analysis and simple regressions

It is well known that, in the term structure literature, three factors are sufficient to explain a large amount of movements of the nominal yield curve. This is confirmed in Panel A of Table V-3, which shows that 100% (99%) of the variation in the level (changes) of nominal and real yields can be explained by the first three principal components, namely the level, slope and curvature factors. All principal components are linear combinations of the included yields. Figure V-10 and Figure V-11 plot the coefficients (or loadings) of these linear combinations as function of the maturities of the nominal and real yields. The loadings of the first principal component are horizontal, which means that a change in this factor will translate to a parallel shift of the nominal and real yields for all maturities (and therefore, this factor is called the level factor). The loadings of the second principal component is downward sloping for the nominal yields but upward sloping for the real yields. This means the second component rotates the yield curve but in different directions for the nominal and real yields. An ambiguous change in this component will shift the short end of the nominal (real) yield curve upward (downward) but decreases (increases) the long end (and therefore, this factor is conventionally named the slope factor). The loadings of the third principal component of the nominal and real yields are hump-shaped and U-shaped respectively. The hump shape occurs at 1-year and 2-year maturities for the nominal yields, whereas the U-turn for the real yield curve occurs at 6-year maturity. Since a change in the third component would increase (decrease) the bond yields

with intermediate maturities relatively to both the short and long ends, this component is also known as the curvature factor. The three independent principal components can also be classified according to their monthly autocorrelation (persistence). Recall the displayed high persistence of both nominal and real yields in Figure V-8 and Figure V-9, the monthly autocorrelations from each factor are shown in Panel B of Table V-3. Among the three principal components, the level factor is the most persistent with a monthly autocorrelation of 0.988 and 0.962 for nominal and real yields accordingly. This feature, to some extent, validates the populous adoption of single factor (short rate) models in early term structure literature. Traditional short rate models, by and large, only model the high persistence of yield movements. The drawbacks of this approach, however, are clearly attributed to the omission of the slope and curvature effects on the yield curve²⁸.

Panel C and D of Table V-3 report the implied fitting errors for nominal and real yields from the principal component analysis, which are calculated as the difference between actual yields and the principal component model predicted yields. Factor models based on principal components predict the yields are linear combination of fixed time series of the filtered factors. The yield coefficients in the prediction are not imposed by no-arbitrage restrictions. Therefore, this simple factor model provides a natural benchmark for the cross-maturity fit. We compute the mean and standard errors of the absolute value of these fitting errors for level, slope and curvature factors. For nominal (real) yields, the less than 9.5 (0.3) basis points fitting error for all maturities suggests that this principal component based factor model performs extremely well, not only in explaining the variance in yields, but also in fitting the evolution of yield dynamics.

²⁸ By incorporating nonlinearities in short rate model, the drawbacks may be substantially mitigated, e.g. see Ait-Sahalia (1996) and Bansal and Zhou (2002).

Table V-3: Principal component analysis

Panel A: Variations in level and changes of nominal and real yields explained by the first three principal components (Level, Slope and Curvature)									Panel B: Autocorrelation from each factor		
	Nominal yields				Real yields				NominalReal		
	Level	Changes			Level	Changes					
Level	0.967	0.787			0.976	0.967			0.988	0.962	
Slope	0.996	0.929			1.000	0.999			0.954	0.927	
Curvature	1.000	0.990			1.000	1.000			0.762	0.894	
Panel C: Absolute value of fitting errors for nominal yields											
Maturity	3	12	24	36	48	60	72	84	96	108	120
Mean	0.039	0.095	0.026	0.038	0.047	0.042	0.030	0.016	0.015	0.032	0.050
S.E.	0.037	0.094	0.023	0.037	0.045	0.039	0.026	0.013	0.015	0.030	0.045
Panel D: Absolute value of fitting errors for real yields											
Mean					0.001	0.003	0.002	0.002	0.003	0.002	0.002
S.E.					0.001	0.002	0.002	0.002	0.002	0.001	0.002

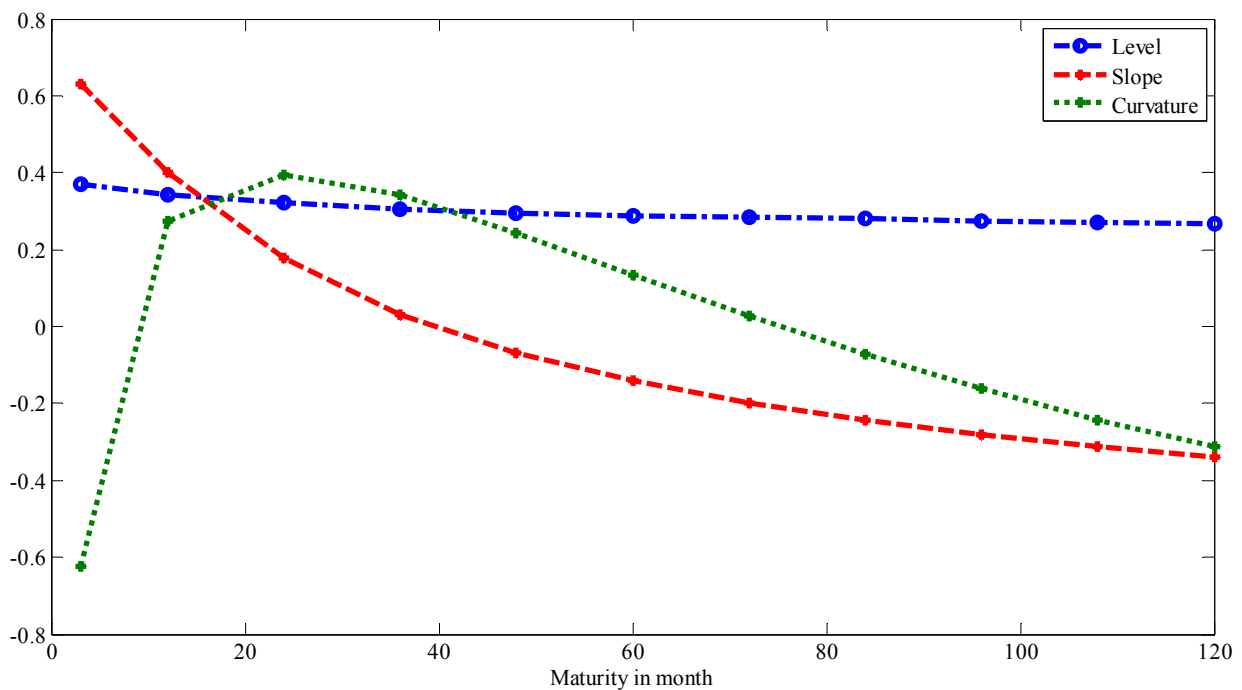
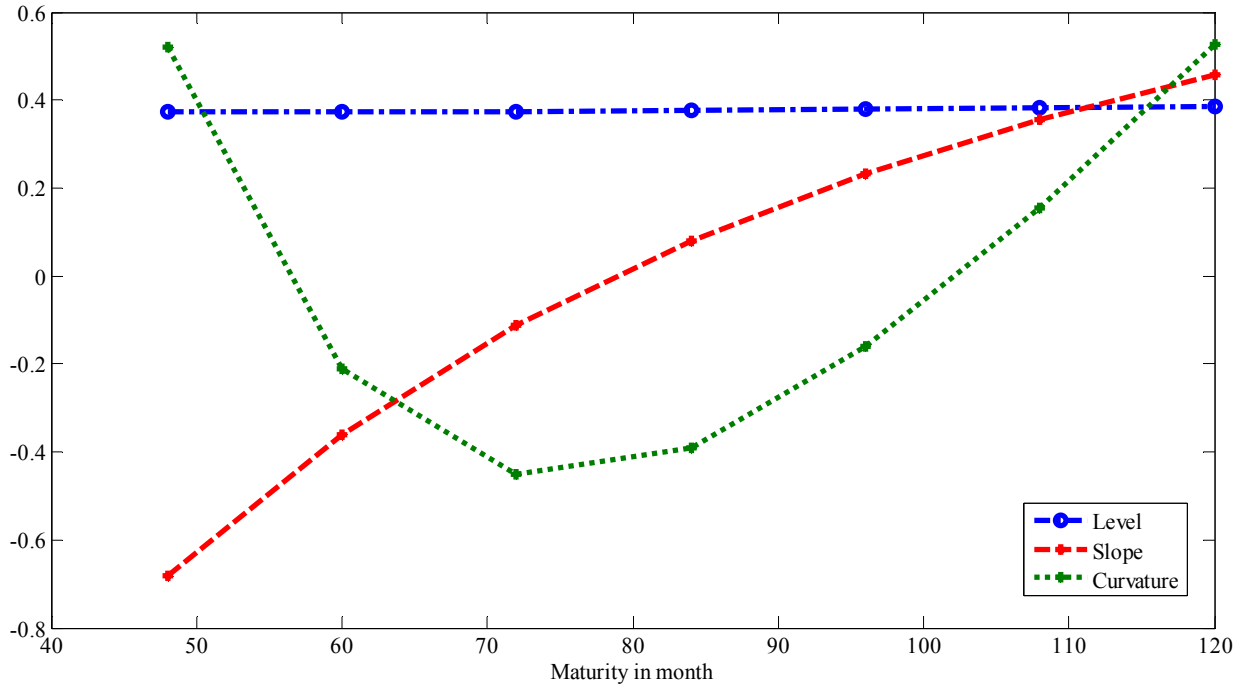
Figure V-10: Factor loadings of nominal yields on principal components

Figure V-11: Factor loadings of real yields on principal components



Given the filtered principal components, we now present some simple regression analysis to test if there is a liquidity factor in the UK inflation index-linked gilts market that needs our attention. As discussed in D' Amico, *et al.* (2010), if the inflation index-linked yield is a good measure of the underlying real yield, the break-even inflation rate, which is the difference between the nominal and index-linked yield, is simply the sum of the expected inflation and inflation risk premia that are also embedded in the nominal yield. Therefore, the projection of the break-even inflation rate onto nominal yield curve factors will achieve a high R^2 . On the contrary, if there are any variations in the index-linked yields that cannot be explained by those nominal yield curve factors (e.g. the liquidity factor), the regression will result a low R^2 . In equation (V.24)²⁹, we show the results of regressing the 10-year break-even inflation rate on the three filtered principal components of nominal yield curve and a constant.

$$BEI_t^{(10\text{-year})} = 4.085 + 0.160 \times Level - 0.185 \times Slope - 0.398 \times Curvature \quad (V.24)$$

(0.052)
(0.007)
(0.032)
(0.078)

The adjusted R^2 from this regression is 0.924, suggesting there is a small portion of variations in the 10-year break-even inflation rate that cannot be accounted by the principal components underlying the nominal yield curve. However, this result, compared with that in D' Amico, *et al.*

²⁹ The standard errors shown in parentheses are Newey-West HAC standard errors calculated with 5 lags.

(2010) on the US TIPS market, shows the liquidity problem in the UK index-linked gilts market is not prominent.

V.6 Estimation

The preliminary data analysis in sections V.4 and V.5 suggest a low dimension of independent factors (at least three) would be sufficient to describe the dynamics of the UK nominal and real yields, given the sample period we studied in this paper. In this section, we discuss relevant estimation issues of the essentially affine model described in section V.3.

Following Ang and Piazzesi (2003), Ang, *et al.* (2008b), Hordahl and Tristani (2007) and many others, we estimate the model using maximum likelihood method based on Chen and Scott (1993). To implement this method, we assume there are three nominal yields that are measured without errors, denoted as y_t^{ne} ; and the rest nominal and real yields are measured with errors, denoted as y_t^e . The reason we need to assume three yields are measured without errors is that we need to use them to back out the three latent factors that drives the nominal and real yield curves. In estimation, we assume the nominal yields with maturities 2-year, 4-year and 9-year are measured without errors. We also tried many other combinations of assumed perfectly measured yields, the results are similar. In comparing with the Kalman filter estimation method used in section III.4.2 of Chapter III, the advantage of using Chen and Scott (1993)'s maximum likelihood estimation method is that it achieves faster convergence in finding global maximum. The drawback, however, is also clear that econometrician has to choose arbitrarily the perfectly measured yields.

To back out the three latent factors, notice the three nominal yields that are assumed to be measured without errors can be written as

$$y_t^{ne} = a^{ne} + b^{ne} X_t, \quad (V.25)$$

where a^{ne} and b^{ne} are the coefficients corresponding to the yields measured without errors given the parameter vector $\Psi = \{\Phi, \Sigma, \alpha, \beta, \lambda_0, \lambda_1\}$ is known³⁰. The latent factors X_t is then obtained by inversion of (V.25), such that the factors are linear in yields as

$$X_t = \frac{y_t^{ne} - a^{ne}}{b^{ne}}. \quad (V.26)$$

³⁰ In backing out the latent factors, notice we do not require the parameters in inflation equation (V.9) are known.

Using the obtained latent factors, the rest of the nominal and real yields that are imperfectly measured can be written as

$$y_t^e = a^e + b'^e X_t + u_t, \quad (\text{V.27})$$

where a^e and b^e denote the coefficients of factors corresponding to imperfectly measured yields and u_t represents the measurement errors that are assumed to be independent and normally distributed with constant variance.

The likelihood value function, constructed in the method of Chen and Scott (1993), consists of two components: the conditional density function of the perfectly measured yields denoted as $f_y(y_t^{ne} | y_{t-1}^{ne})$, and the density function for the measurement errors in those imperfectly measured nominal and real yields denoted as $f_u(u_t)$. Since there is a one-to-one density transformation from the perfectly measured yields to the latent factors (as in (V.26)), the conditional density function of the perfectly measured yields can be written as a Jacobian transformed density function of latent factors as

$$f_y(y_t^{ne} | y_{t-1}^{ne}) = \frac{f_x(X_t | X_{t-1})}{|\det(J)|}, \quad (\text{V.28})$$

where J is the Jacobian transformation matrix (the b'^{ne} in (V.26)) and $f_x(X_t | X_{t-1})$ is the conditional density function of the latent factors. Therefore, the log-likelihood function can be constructed as

$$\begin{aligned} L(\Psi) &= \sum_{t=2}^T \left(-\ln(|\det(J)|) + f_x(X_t | X_{t-1}) + \ln(f_u(u_t)) \right) \quad (\text{V.29}) \\ &= -(T-1) \ln(|\det(J)|) - \frac{T-1}{2} \ln(|\det(\Sigma \Sigma')|) \\ &\quad - \frac{1}{2} \sum_{t=2}^T \frac{(X_t - \Phi X_{t-1})' (X_t - \Phi X_{t-1})}{\Sigma \Sigma'} \\ &\quad - \frac{T-1}{2} \sum_{i=1}^{15} \ln(\sigma_{u,i}^2) - \frac{1}{2} \sum_{t=2}^T \sum_{i=1}^{15} \frac{u_{t,i}^2}{\sigma_{u,i}^2} \end{aligned}$$

where $\Psi = \{\Phi, \Sigma, \alpha, \beta, \lambda_0, \lambda_1, \delta_0, \delta_1\}$ is the parameter vector that maximizes (V.29)³¹.

³¹ Since we have 8 out of 11 nominal yields and 7 out of 7 real yields that, in assumption, are measured with errors; the density function $f_u(u_t)$ consists of 15 measurement error densities.

To implement the maximum likelihood estimation exercise in an affine term structure model, one often encounters a practical problem that the gradient-based optimization methods typically converge very slowly. This is due to the persistence of the factors, which make the diagonal coefficients of Φ in (V.4) near unity. When these autoregressive coefficients are close to one, a tiny optimization step around some fair values of them would still be too far away from some reasonable values of the VAR(1)'s drifts (if they are specified in the model). The likelihood function, thereby, will be essentially flat in the drifts but steep in autoregressive coefficients. In order to circumvent this problem, as in (V.4), we omit the drift terms in the VAR(1) system of latent factors. In addition, we set the initial values of these autoregressive coefficients to 0.98, 0.95, and 0.76, which are the autocorrelation coefficients for level, slope and curvature factors reported in Panel B of Table V-3. The starting values for latent factors' volatilities are set at 0.0001, as suggested in Risa (2001). To identify the short rate equation parameters (α and β) and the expected inflation (π_t), we normalise α to equal the mean of the 3-month short rate and set β to be a vector of ones, while let δ_0 and δ_1 be determined by the data. By doing this, we implicitly allow the real yield and the model implied expected inflation to be correlated, which is well-known as the Mundell-Tobin effect. Finally, for the starting values of the market price of risk coefficients, we initiate them with zeros values, which corresponds to the risk neutral case. The optimization is undertaken by Matlab's optimization routine, with at least 20 different sets of initial values considered in order to obtain a global maximum. The Matlab codes are provided in Appendix V.

V.7 Empirical results

The estimation results of the essentially affine term structure of the nominal and real yields are reported in Table V-4. The estimated three autoregressive coefficients, which ranges from 0.9654 to 0.9928, indicate that the three factors are highly persistent. The cross factor coefficients (the off-diagonal elements in Φ) are statistically highly significant. One unit change in the first latent factor in the previous period will cause 0.024 unit decrease in the second factor and a 0.6139 units increase in the third factor, one period later. Meanwhile, the second factor in previous period is influencing positively on the third factor in later period. The coefficients of market price of risk are all different from zeros, which implies a rejection of risk-neutral case and suggests a time-varying price of risk.

Table V-4: Parameter estimation results

Parameters	Estimates	Standard Errors
$\phi_{1,1}$	0.9928	0.0063
$\phi_{2,1}$	-0.0240	0.0198
$\phi_{2,2}$	0.9790	0.0120
$\phi_{3,1}$	0.6139	0.2418
$\phi_{3,2}$	0.2061	0.0354
$\phi_{3,3}$	0.9654	0.0164
σ_1	2.503E-05	4.577E-08
σ_2	1.122E-04	1.235E-11
σ_3	3.908E-04	1.903E-05
$\lambda_{0,1}$	0.0020	0.0179
$\lambda_{0,2}$	0.0181	0.0250
$\lambda_{0,3}$	0.0040	0.1187
$\lambda_{1,1}$	-41.8338	89.2462
$\lambda_{1,2}$	0.3638	26.1775
$\lambda_{1,3}$	-1.5889	7.5611
$\lambda_{2,1}$	600.0342	186.1279
$\lambda_{2,2}$	-69.4655	104.2516
$\lambda_{2,3}$	-8.3973	5.8887
$\lambda_{3,1}$	329.6110	608.5285
$\lambda_{3,2}$	1.7413	44.2010
$\lambda_{3,3}$	105.6900	30.5174
δ_0	0.0032	1.907E-05
$\delta_{1,1}$	-6.2785	1.5534
$\delta_{1,2}$	-0.8165	0.3176
$\delta_{1,3}$	0.9574	0.0421
Log-likelihood value	48178.54	

The three filtered latent factors are plotted in Figure V-12. In general, the last two factors display a declining trend over the sample period, which is in accordance with the declining trend of the UK's nominal and real rates exhibited in Figure V-1 and Figure V-3.

The model fitted nominal and real yields are plotted in Figure V-13 to Figure V-16. Since the latent factors are backed out from three nominal yields, the model fits remarkably well on nominal yields both across maturities and over time but less promising on real yields.

Figure V-12: Filtered three unobserved factors

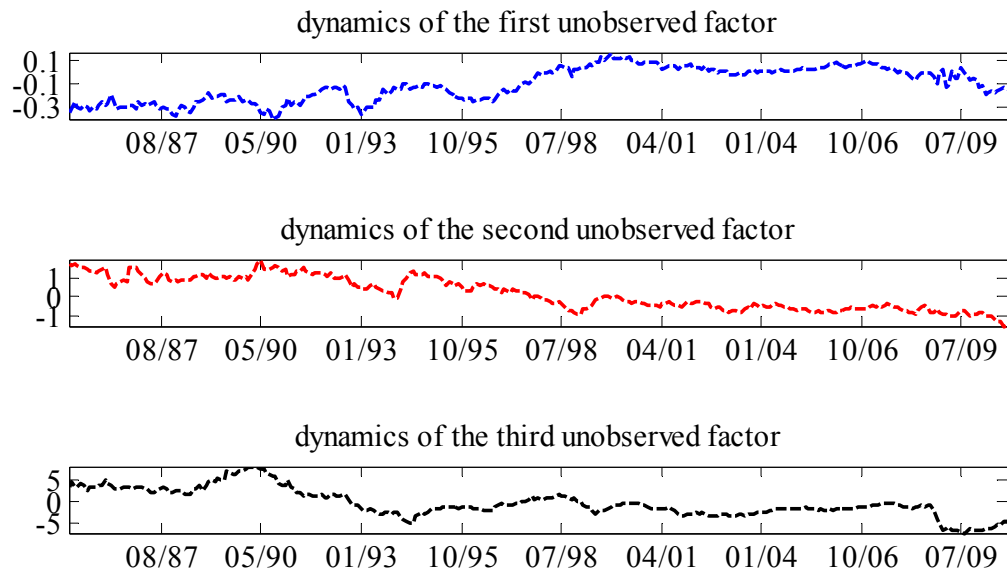


Figure V-13: Observed and fitted nominal yields (3-month, 1-year, 3-year and 5-year)

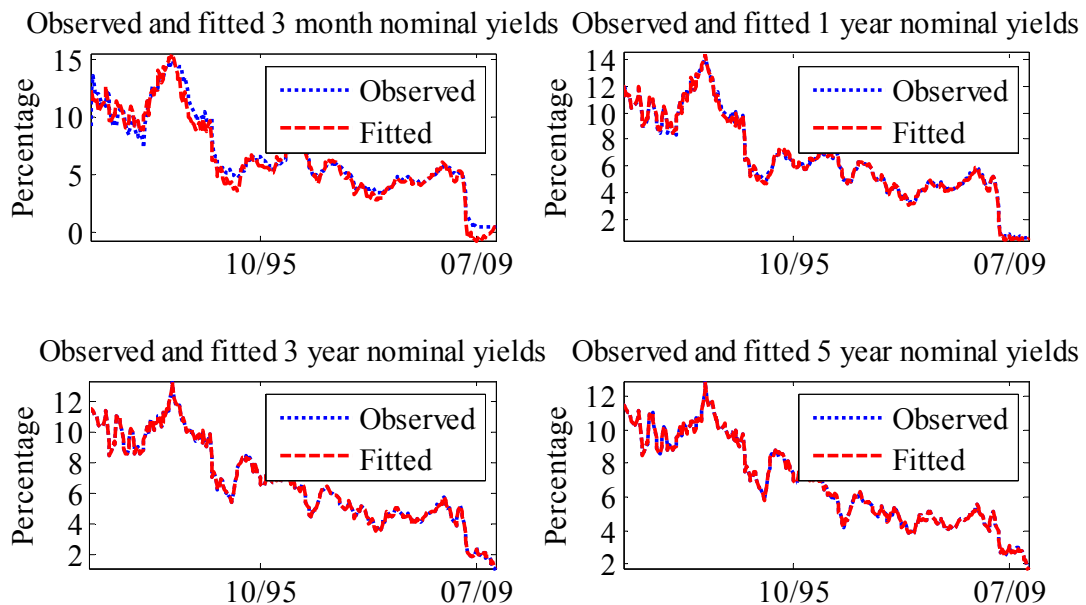


Figure V-14: Observed and fitted nominal yields (6-year, 7-year, 8-year and 10-year)

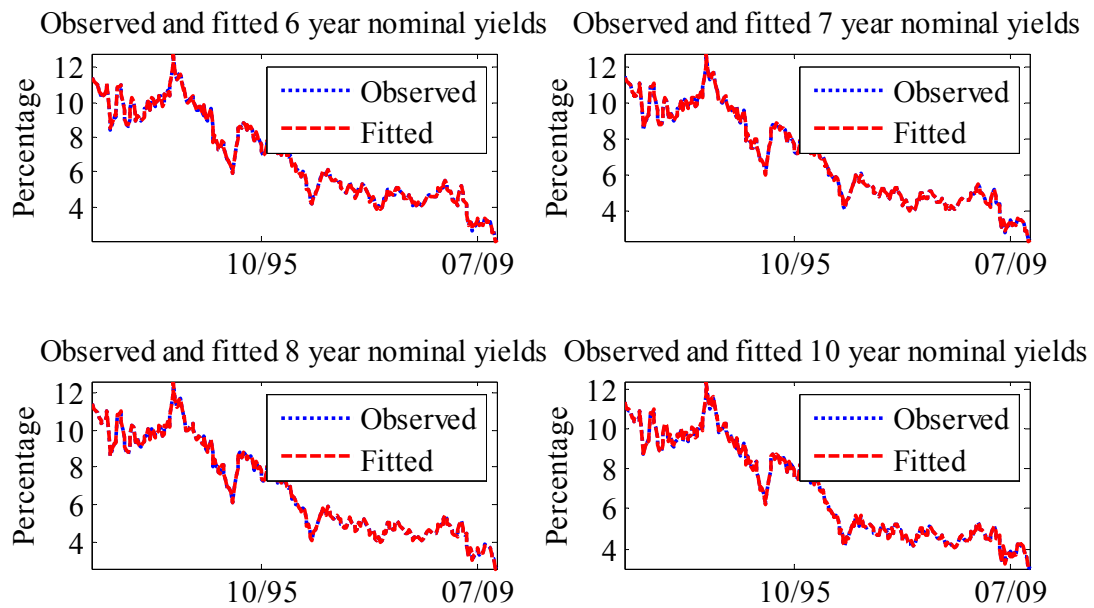


Figure V-15: Observed and fitted real yields (4-year, 5-year, 6-year and 7-year)

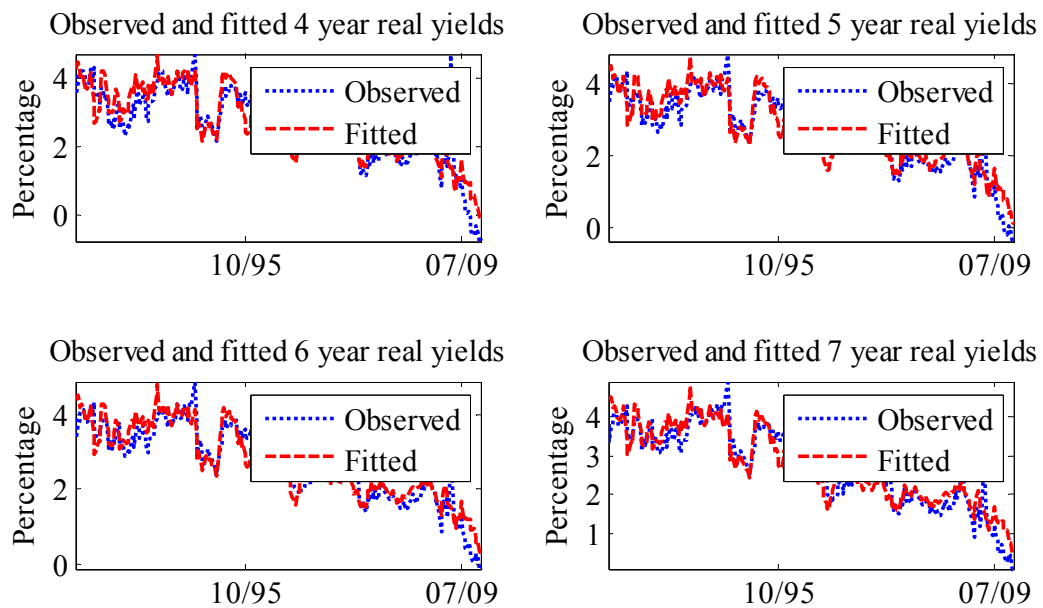
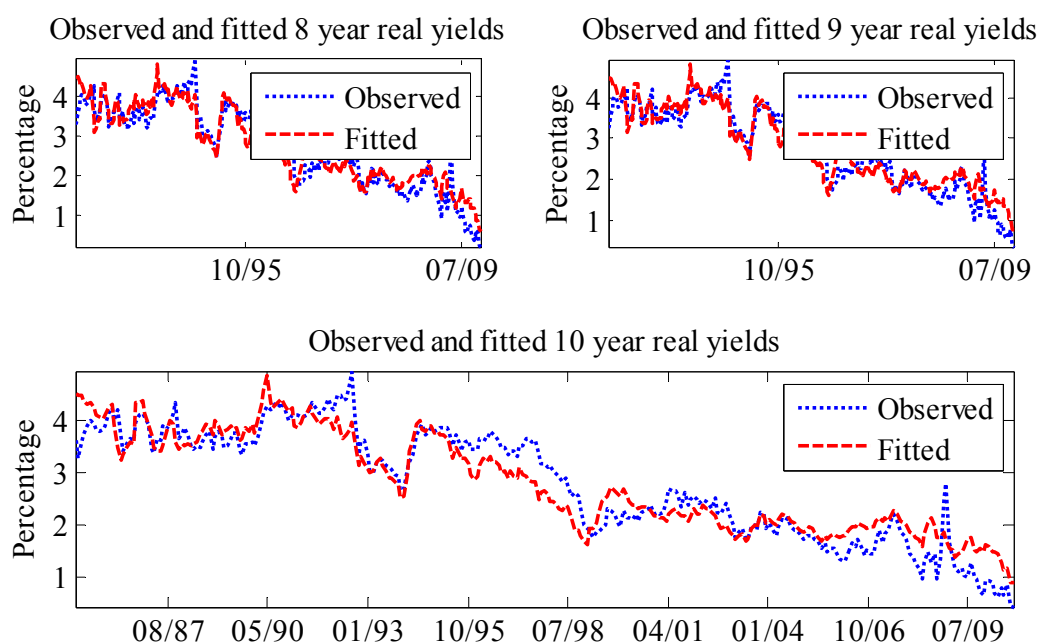


Figure V-16: Observed and fitted real yields (8-year, 9-year and 10-year)



The estimated nominal term premia, defined in equation (V.20), are depicted in Figure V-17 and Figure V-18 for yields with 5-year and 10-year maturities. Notice the dotted lines represent the difference between the estimated nominal yields and term premia, which is the implied pure expected nominal yields³². From these figures, we can see the nominal term premia varies significantly over time and showed a significant presence in early sample periods. As we discussed in Section V.1.1, the mid-80s has seen a recovery from the second oil crisis where the UK's economy was stable during that time period. The 5-year and 10-year nominal term premia declined gradually from roughly 100 and 200 basis points to 0 and 100 basis points respectively in the late 80s as a result of the booming GDP growth. The later inflation spikes in 1990 and early 90s' recession, however, once again brought up the nominal term premia until when the British government joined the ERM. The premia (10-year) thereafter decreased mildly till the next spike when the British government exit the ERM in September 1992. When the first Britain's inflation target (1%-4%) was brought out subsequently in October, the nominal term premia started to decline. A further decline was after the Bank of England's independence in 1997, which is regarded as a signal of Britain's central bank's commitment in anchoring inflation rate. In fact, the 5-year and 10-year nominal term premia were kept well below 100 basis points after 1997 until the broke out of recent financial crisis.

³² In pure expectation theory of the term structure of interest rates, the term premia, by assumption, is zero.

Figure V-17: 5-year nominal yields and term premia

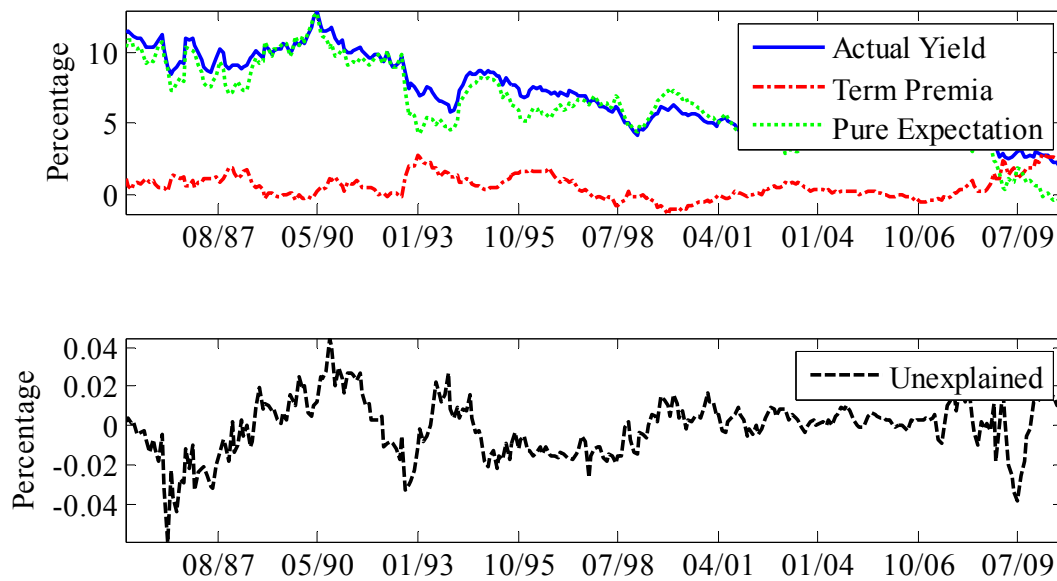
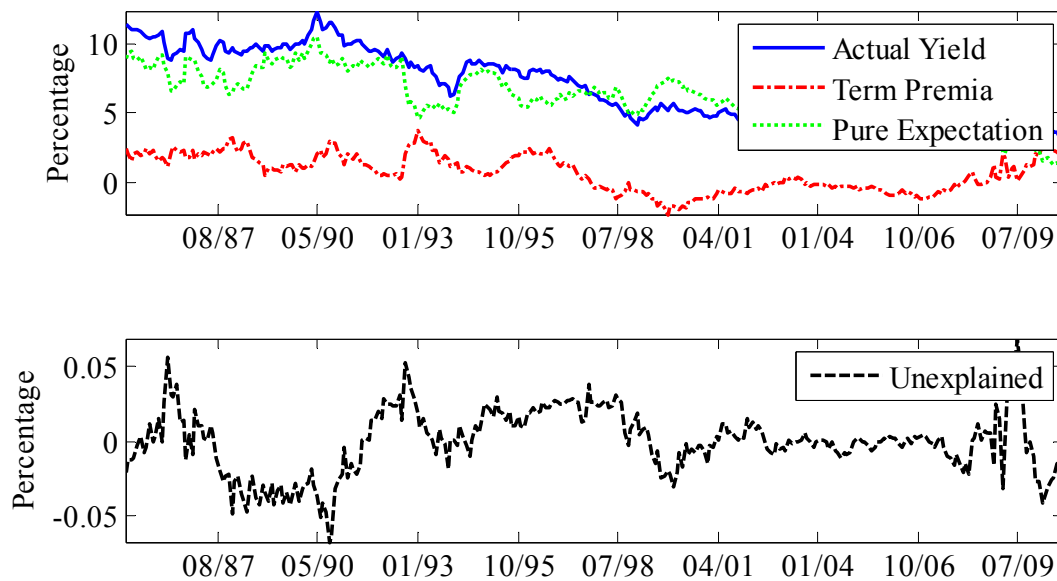


Figure V-18: 10-year nominal yields and term premia



The real term premia for maturities of 5-year and 10-year, as presented in Figure V-19 and Figure V-20, show the similar declining trend over the course of the sample period. Notice the sharp decline after the independence of the Bank of England in 1997, which essentially brought the real term premia into a negative regime thereafter. The recent financial market

turmoil, which lifted up investors' uncertainties about the future real economy, has seen a sharp rise in the real term premia.

Figure V-19: 5-year real yields and term premia

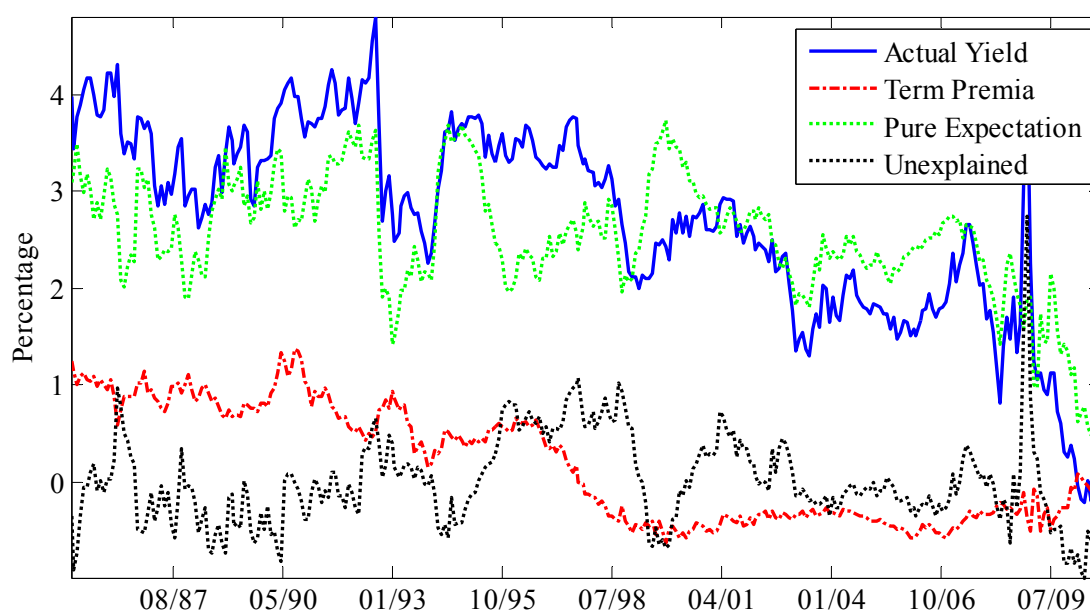
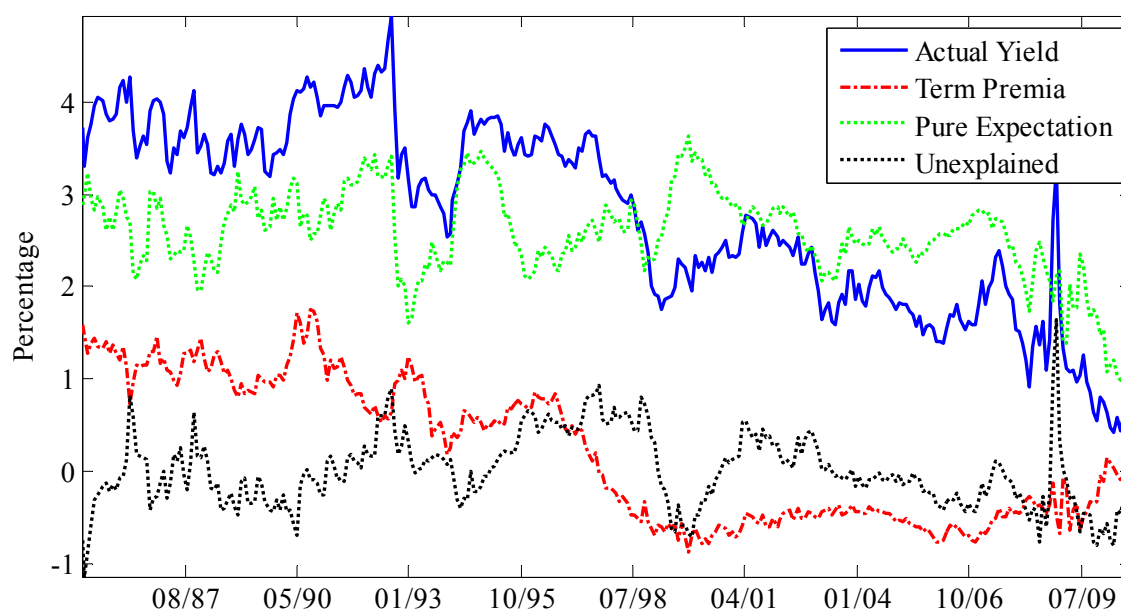


Figure V-20: 10-year real yields and term premia

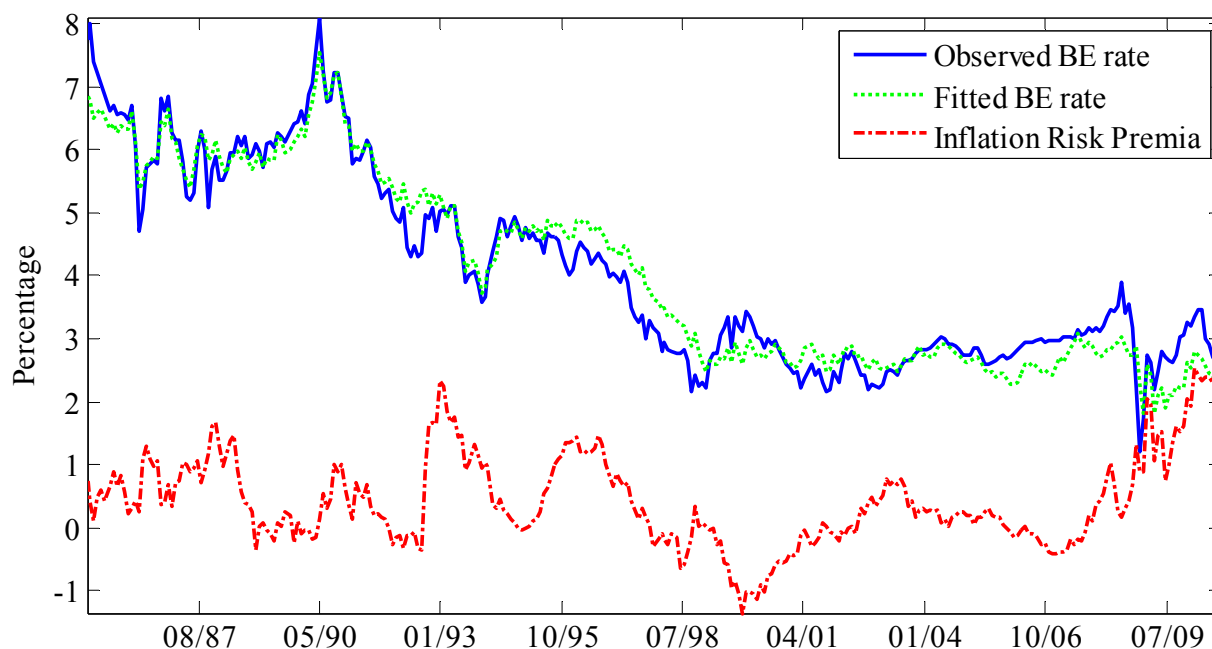


The strong resemblance in falling for the nominal and real term premia in the period after the independence of the Bank of England is suggested by both declines in inflation risk and real risk. Figure V-21 plots the 10-year break-even rate and inflation risk premia. Our results is similar to Remolona, *et al.* (1998) and Risa (2001), who find a moderate 100 basis points before 1997 and subsequently lower after that. The next wave of increasing in the

inflation risk premia starts from the beginning of the new millennium till recently arrives at a level of 200 basis points.

Although the independence of the Bank of England contributes to the sharp fall of the real term premia after 1997, the degree of influence, as argued in Joyce, *et al.* (2010), is unlikely to be large. One important factor that is likely to influence the real term premia largely is the introduction of the Minimum Funding Requirement, which protects the solvency of pension funds, becomes effective in April 1997. This regulatory reform led to an increase in the demand for inflation index-linked gilts, which effectively compressed the real term premia. Other factors, like the Asia financial crisis in 1997 and the LTCM crisis in 1998, also caused large demand in the government bonds due to the "flight-to-quality".

Figure V-21: 10-year break-even rate and inflation risk premia



The term structure of the average yield curve, the volatility and the autocorrelation of nominal, real and inflation risk term premia are plotted in Figure V-22 to Figure V-25. Generally, the average and volatility term structure of the nominal, real and inflation risk term premia are all sloping upwards. In contrast, although the autocorrelation decreases with maturity, the absolute value for all maturities is very high.

Figure V-22: Term structure of the average nominal and real term premia with ± 2 S.E. bounds

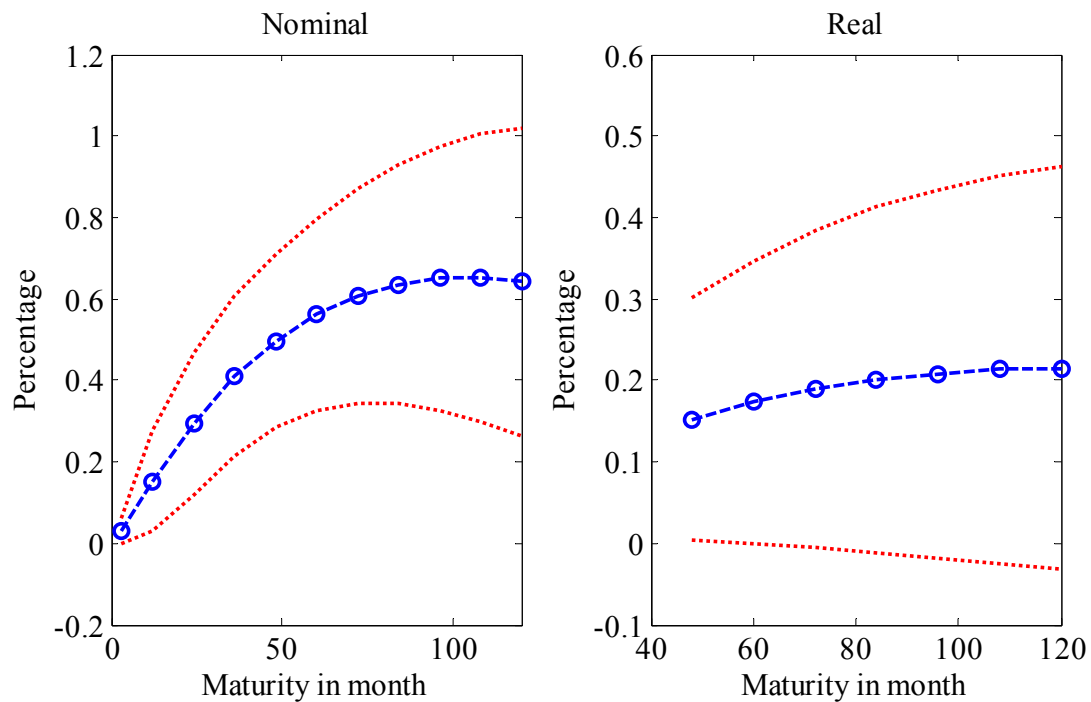


Figure V-23: Term structure of the volatility of nominal and real term premia with ± 2 S.E. bounds

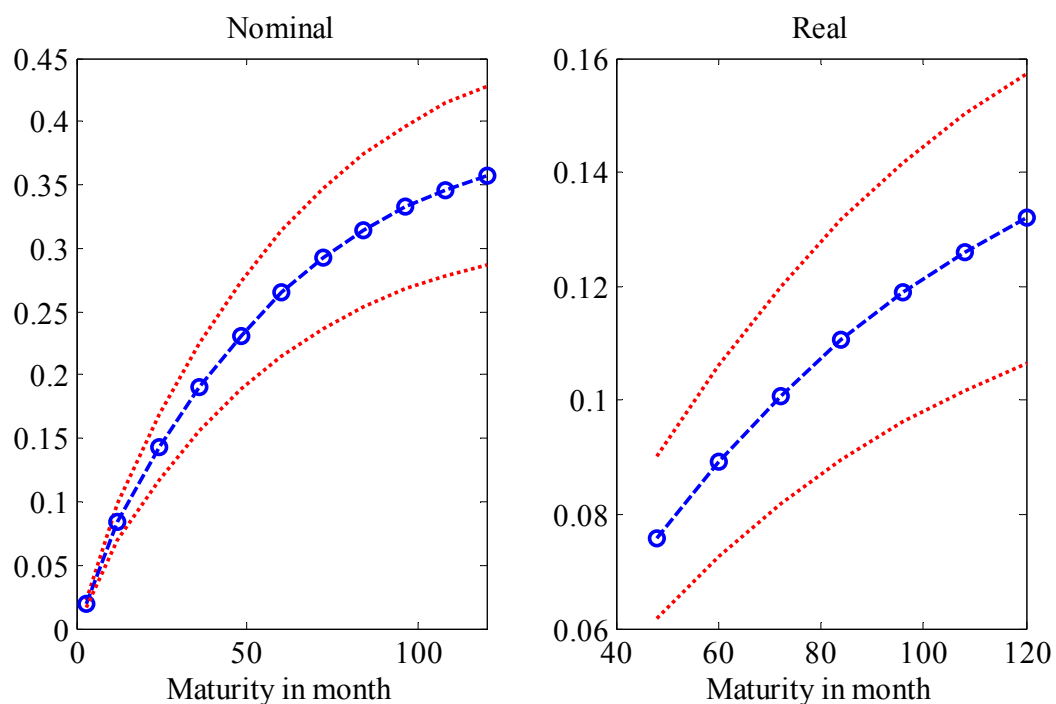


Figure V-24: Term structure of the autocorrelation of nominal and real term premia with ± 2 S.E. bounds

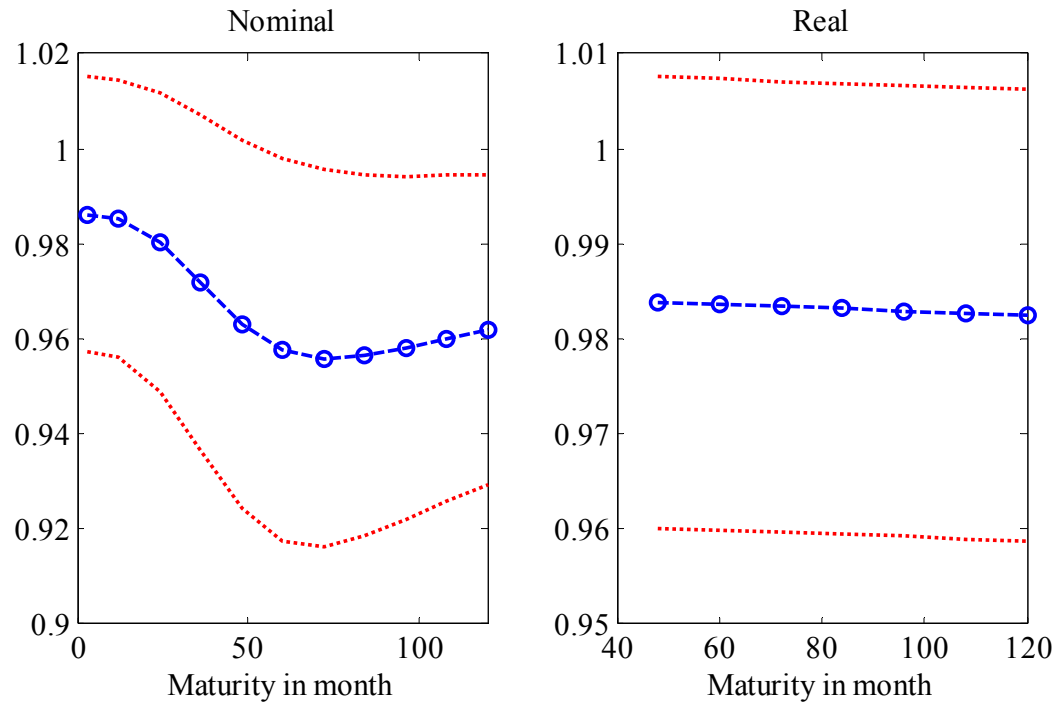
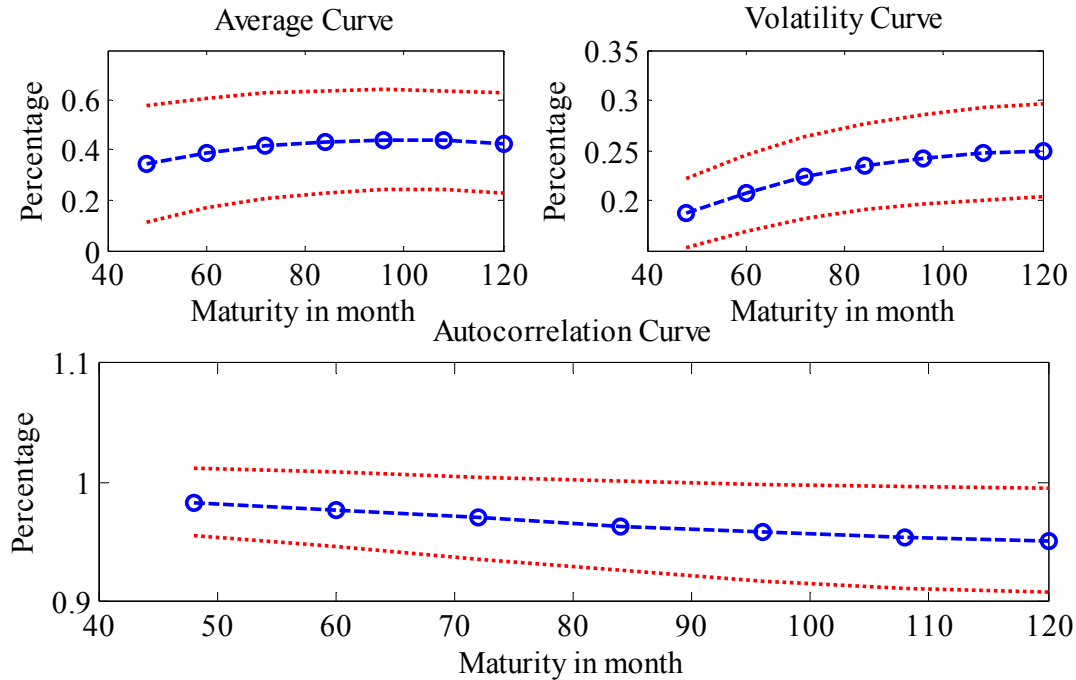
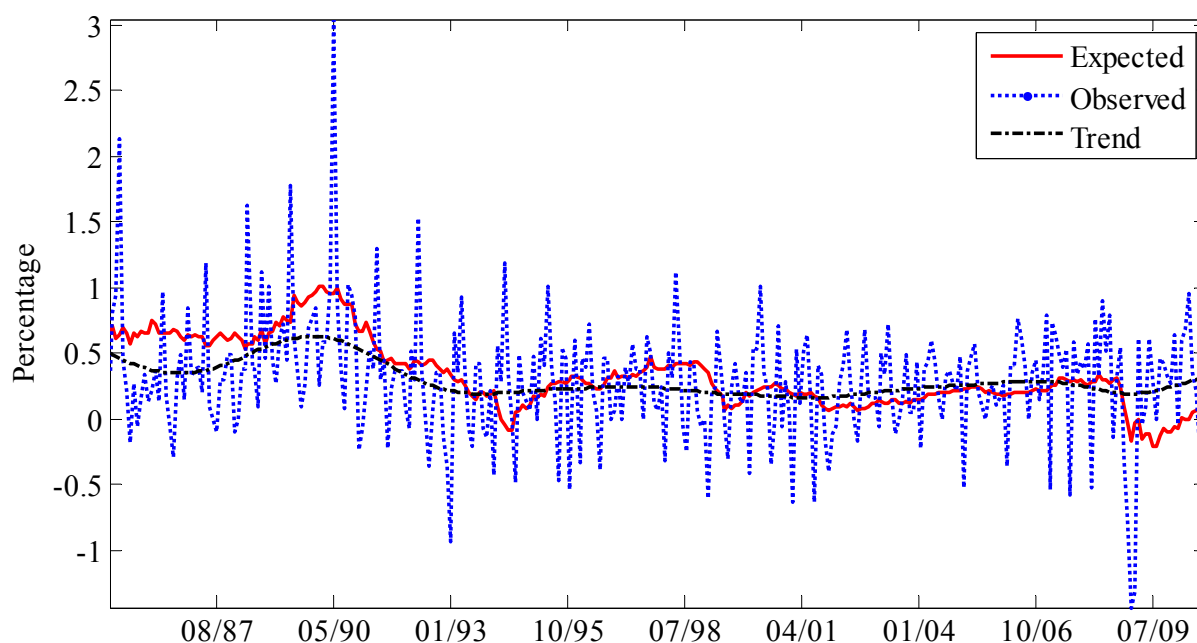


Figure V-25: Term structure of the inflation risk premia (average, volatility and autocorrelation curves) with ± 2 S.E. bounds



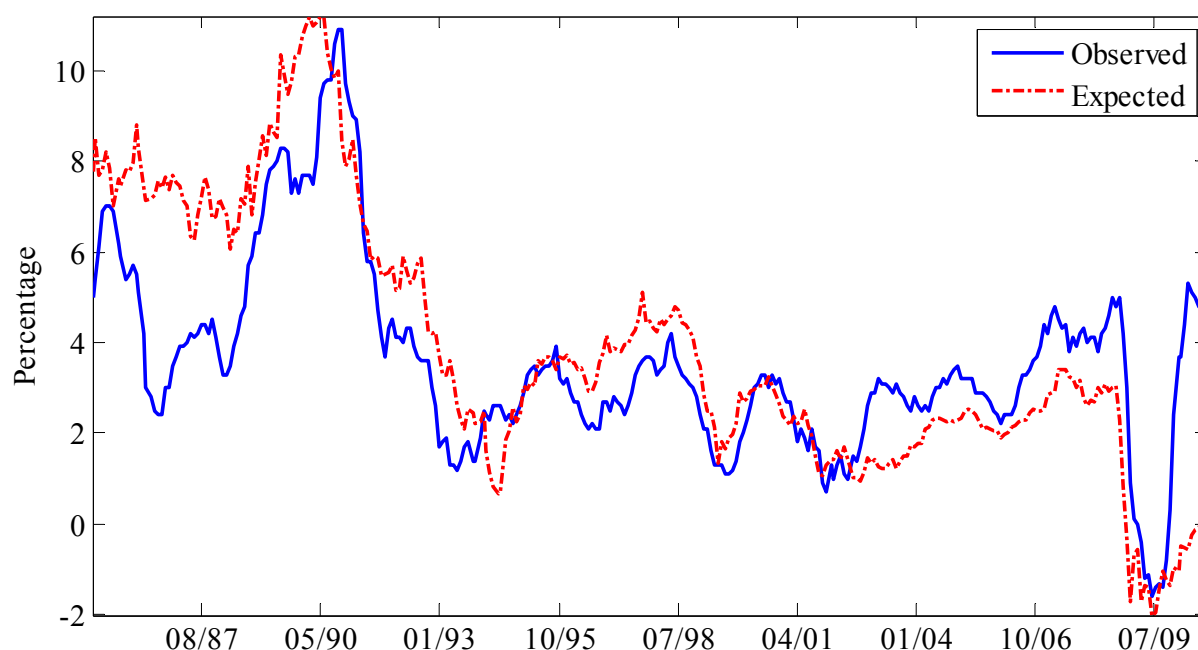
In Figure V-26 and Figure V-27, we compare the observed and expected 1-month and annual inflation rates³³. The black dotted line in Figure V-26 is the trend of the observed 1-month RPI inflation rate, calculated using HP filter. The smooth expected month-on-month inflation rate rather behaves like the trend extracted from the observed series. On average, the expected series is above the trend before the year of 2000, but remains below it in the subsequent periods. Comparatively, the expected annual inflation rate also overstates the actual inflation rate before the new millennium and understates it in the following periods. This may suggest that the Bank of England's monetary policy is eventually quite creditable after it became independence in 1997. Most astonishingly, the recent sharp decline of the expected inflation rates (which even lead to a deflation) may lend support to the standing ground of the central bank to keep interest rates at historically low level. The background of the decline in expected inflation during this time period is that the global economy suffered a sharp and synchronised downturn. Along with business and households' sentiment in the UK has deteriorated markedly, GDP contracted sharply in the fourth quarter of 2008. Business surveys pointed to similar reduction in output in early 2009 (see Inflation Report, February 2009, Bank of England). Some economists even argue about the possibility of core inflation becomes negative in 2009 (see BBC news: <http://news.bbc.co.uk/1/hi/business/7893873.stm>).

Figure V-26: The observed and expected 1 month inflation rates



³³ The observed month-on-month and year-on-year RPI inflation rates are obtained from the website of the UK Debt Management Office.

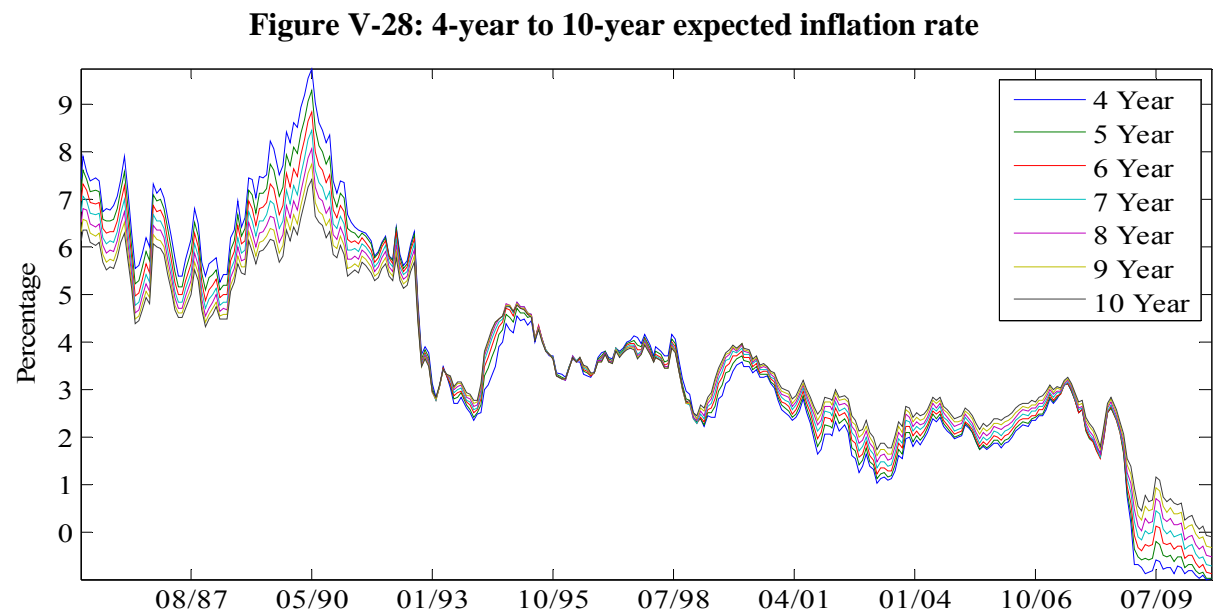
Figure V-27: Observed and expected annual inflation rates



From Figure V-27, we see the expected annual inflation rate has declined significantly from the peak in 1991 to around 2% for decade long after 1992. This is a particularly interesting period for the UK as its economy has undergone significant statutory changes in the early 90s: The UK joined the ERM in October 1990, and forced to leave on the September 16th 1992, following a massive wave of currency speculation, after which a regime of inflation targeting was then established three weeks after the suspension of ERM membership; in 1997 the Bank of England became independent and acquired the responsibility for setting interest rates to meet inflation target. Along with the regime changes in monetary policy, the post-1992 period has seen the volatilities of several macroeconomic indicators have reached the lowest level since the collapse of Bretton Woods (see Benati (2004)). Although there are still many debates about whether the improved monetary policymaking has been largely responsible for the lower volatilities in macroeconomic indicators, it has been widely recognised that monetary policy may have been important in reducing volatilities to the extent which policy changes have resulted in lower and more stable inflation (see Summers (2005)). By achieving low and stable inflation, this translates into low and stable expected inflation, which reinforces a favourable environment for economic activity that removes sources of uncertainties that might cloud investors' investment decisions.

In Figure V-28, we plot the 4-year to 10-year expected inflation rates. The shorter maturity inflation rates are generally substantially higher than the ones with longer maturities

in the early period of our sample. But the situation is reversed after the year 2000. Not only the level of the expected inflation has decreased continuously throughout the sample period, the gap between the long run and the short run expected inflation rates has also decreased since 1992, when the first Britain's inflation target (1%-4%) was adopted. The recent financial crisis, which brought down the inflation expectation, reflects largely on investors' worries about deeper future recessions.



One thing that needs to be taken with caution is that, even the expected inflation rates in recent periods are close to zero, the 10-year inflation risk premia as plotted in Figure V-21 is relatively high since the inception of the recent financial crisis. Using exactly the same logic as in V.5, we regress the 10-year break-even rates on 3-month, 5-year and 8-year nominal yields for the periods January 1985 to December 2007 and January 2008 to September 2010. The regression results are presented in Table V-5. The adjusted R-square statistics in the second subsample is merely 39% compared to the 95% of the explanatory power of the three nominal yields in the pre-crisis period, which suggests that there are other factors, other than the expected inflation and inflation risk premia, contained in the nominal yields during the financial crisis period. One possible explanation on this is that the inflation risk premia estimated for this particular sample period may contain a large amount of liquidity premia.

Table V-5: Projection of the break-even rates on nominal yields within different sample periods

	January 1985 to December 2007	January 2008 to September 2010
--	-------------------------------	--------------------------------

	Coefficients	S.E.	Coefficients	S.E.
Constant	-0.3027	0.0685	-0.2759	0.8810
$y_t^{(3month)}$	0.0918	0.0183	0.1533	0.1082
$y_t^{(5year)}$	-0.5000	0.1137	-1.3244	0.5167
$y_t^{(8year)}$	1.0300	0.1010	1.8904	0.5576
Adjusted R-square	0.9523		0.3877	

V.8 The relation between stock returns and expected inflation

While macroeconomists are interested in whether monetary policy has any effect on the real economy activity (including the real return of stocks), financial economists often focus on the question if stock market is a good hedge against inflation. Over the past 30 years, the puzzle of negative (or non-existent) relation between stock returns and inflation has sparked a significant amount of research into the financial economists' question of whether stocks are good hedge against expected inflation. Following the seminal work of Bodie (1976), the relation between stock returns and inflation has been extensively tested in the context of Fisher's hypothesis of interest (Fisher (1930)), which states that nominal expected return consists of expected real interest rate and the expected inflation rate. The extended Fisher hypothesis perceives a positive one-to-one relation between stock returns and inflation. However, the well documented empirical findings (e.g. Fama and Schwert (1977), Gultekin (1983), Amihud (1996), Barnes (1999) and many others) suggest this relation to be negative, only a few have found a positive relation (e.g. Luintel and Paudyal (2006)), especially in a long-run cointegration relationship.

Virtually all previous studies, except Barnes (1999), Boyd, Levine and Smith (2001), and Kim (2003), that we are aware of, tested the stock return and inflation relation in a linear fashion, that is assuming the response of stock return to the inflation rate is constant through high and low inflation periods. However, the assumption of symmetric response may lead to misspecification of the true underlying relation between stock returns and inflation rates, if there are any. In this section, we improve on previous studies of relation between the stock returns and inflation rates by employing a nonlinear Smooth Transition Vector Autoregressive (STVAR) model.

By doing so, we assume the underlying relation between stock returns and inflation rates varies over time. In particular, we postulate that the response of stock returns to inflation rates (or the reverse) is regime dependent. Such dependence is accounted for by considering regime changes in the form of smooth transition. This will capture the asymmetric response of

stock returns to the dynamics of inflation rates over time and indicate whether the relation between the two variables is linear or nonlinear³⁴. Notably, Kim (2003) find an asymmetric manner of causality based on quarterly data for German stock returns and inflation rates by using a nonlinear regression analysis, suggesting that the potential nonlinear relation between stock return and inflation rate should be considered with care. Barnes (1999) detects a threshold effect in the relation between stock returns and inflation rates using quarterly and monthly data for 39 different countries. In a TAR formation, the author finds that both the negative stock return and inflation relation in "low-average-inflation" countries and the positive relation in "high-average-inflation" countries exacerbate in a second regime. Similar findings are also reported in Boyd, *et al.* (2001), where they find that increases in inflation is unrelated to increases in nominal stock returns if the annual inflation rate is less than 15%, whilst an almost one-to-one response on stock returns to inflation increases exists at higher inflation rates. This finding suggests that stocks are good hedge against inflation only if the inflation rates are high enough and reach a certain threshold value.

In addition to the methodology advances of utilizing a nonlinear STVAR model, we also differ from previous studies in choosing proxies for expected inflation rates. Since expected inflation is more relevant in investors' asset allocation decisions, yet it is generally unobservable, previous studies on the relation between stock returns and inflation rates have diversely attempted to generate and use different proxies for expected inflation, which includes: (1) contemporaneous realized inflation rates as proxies assuming rational expectation (e.g. Gultekin (1983), Boudoukh and Richardson (1993) and many others), (2) observed expected inflation obtained from professional surveys (e.g. Sharpe (2002) and Schmeling and Schrimpf (2010)), (3) expected inflation calculated from lagged Treasury bill rate assuming constant expected real return in Fisher's equation (e.g. Fama (1981)). In this study, we use a new proxy for the expected inflation - the in-sample/ex post expected inflation rate, which is estimated in the previous section.

A roadmap is needed here. In the following subsections, we first outline the basic STVAR model that we use in this study in section V.8.1. In section V.8.2, we describe the specification tests, which include linearity test and model adequacy tests. Finally, in section

³⁴ Some recent empirical works on other topics but using the STVAR model as an analysing framework includes Lekkos and Milas (2004), and Lekkos, Milas and Panagiotidis (2007).

V.8.3, we present the empirical results based on the monthly stock returns from FTSE 100 index and the monthly expected inflation rates obtained previously.

V.8.1 The STVAR model

The model we use to model the system of stock returns and inflation rates are belong to the Smooth Transition Regression (STR) class. The basic STR model formulation with univariate time series has been discussed in section II.1.1. As an extension to the univariate case, the STVAR model allows us to model multivariate time series in a unified framework, which could shed some lights on the interconnection between the dynamics of the two variables through time.

Consider the following STVAR model proposed by Granger and Terasvirta (1993) and van Dijk (1999), and further extended by Camacho (2004):

$$\begin{aligned} y_t &= z_t' \phi_1^y + (z_t' \phi_2^y) G(\gamma^y, c^y; s_t^y) + u_t^y, \\ x_t &= z_t' \phi_1^x + (z_t' \phi_2^x) G(\gamma^x, c^x; s_t^x) + u_t^x, \\ z_t &= [1, y_{t-1}, y_{t-2}, \dots, y_{t-p}, x_{t-1}, x_{t-2}, \dots, x_{t-p}]', \end{aligned} \quad (\text{V.30})$$

where y_t and x_t are monthly stock returns and changes of the monthly ex post expected inflation rate³⁵, ϕ_1^y and ϕ_2^y (ϕ_1^x and ϕ_2^x) are the two distinct sets of parameter corresponding to z_t in two linear paths of y_t (x_t), u_t^y and u_t^x are the two serially uncorrelated errors, which we assume the are both normally distributed with zero means and variances of Ω^y and Ω^x , accordingly. $G(\gamma^y, c^y; s_t^y)$ and $G(\gamma^x, c^x; s_t^x)$ are the two transition functions for y_t and x_t , respectively. By convention, these two transition functions are bounded between zero and one. If $G(\gamma^y, c^y; s_t^y)$ and $G(\gamma^x, c^x; s_t^x)$ are zeros, we have a linear VAR model with parameters $\phi_{1,t}^y$ and $\phi_{1,t}^x$, if, on the other hand, $G(\gamma^y, c^y; s_t^y)$ and $G(\gamma^x, c^x; s_t^x)$ are ones, we have another linear VAR model with parameters $(\phi_1^y + \phi_2^y)$ and $(\phi_1^x + \phi_2^x)$. Therefore, $G(\gamma^y, c^y; s_t^y)$ and $G(\gamma^x, c^x; s_t^x)$ act to locate the model between two linear VAR models from two distinct regimes.

³⁵ Since a unit root has been detected in the ex post expected inflation rate, we take the first difference of the expected inflation rate and use this stationary series in the STVAR modeling instead.

As we discussed on section II.1.1, there are many ways to define a transition function. In this study, we consider two types of the transition function - the logistic transition function and the exponential transition function. In the case of a logistic transition, $G(\gamma^y, c^y; s_t^y)$ and $G(\gamma^x, c^x; s_t^x)$ are two monotonically increasing functions, taking the form of

$$G^i(\gamma^i, c^i; s_t^i) = \frac{1}{1 + \exp[-\gamma^i(s_t^i - c^i)]} \quad (\text{V.31})$$

where γ^i is the transition speed parameter, $i = y, x$, s_t^i is the transition variable and c^i is the transition threshold that determines the point of transition between regimes. The resulting STVAR model with a logistic transition function is called the Logistic STVAR (LSTVAR). In the case of an exponential transition, $G(\gamma^y, c^y; s_t^y)$ and $G(\gamma^x, c^x; s_t^x)$ are defined as

$$G^i(\gamma^i, c^i; s_t^i) = 1 - \exp\left[-\gamma^i(s_t^i - c^i)^2\right] \quad (\text{V.32})$$

where $i = y, x$, and we call the STVAR model with an exponential transition function as the Exponential STVAR (ESTVAR). Note that, the choice of the transition variable (s_t^i) in equations (V.31) and (V.32) can be the lagged values of y and x , or one can choose an exogenous variable other than y and x as the transition variable. Additionally, we follow Camacho (2004) to assume the same type of transition is fitted for each equation in the system, i.e. equations are ruled by only one transition variable ($s_t^y = s_t^x = s_t$). This implies the economy has only one common nonlinear feature.

The economic interpretation for the LSTVAR and ESTVAR model, when applied to the relation between stock returns and inflation rates, is different. For example, assuming the lag structure of the two models is order-1, which means $z_t = [1, y_{t-1}, x_{t-1}]'$, and the transition variable is the lagged values of the expected inflation rate, the system of equations in (V.30) describes the forecasting power of lagged dependent variables on each other as measured by ϕ_j^i , $i = y, x$, $j = 1, 2$. If both γ^i and ϕ_j^i are significant and greater than zero, in LSTVAR, $G^i(\cdot)$ close to zero implies the value of the transition variable (lagged expected inflation rate) is much less than the threshold value c^i . Then the forecasting powers of y_{t-1} and x_{t-1} are mainly determined by ϕ_1^y and ϕ_1^x in regime 1. On the other hand, as $G^i(\cdot)$ approaches unity, it

implies the transition variable (lagged expected inflation rate) is much higher than the threshold value c^i . Then the forecasting power of y_{t-1} and x_{t-1} are determined by $(\phi_1^y + \phi_2^y)$ and $(\phi_1^x + \phi_2^x)$ in regime 2. Therefore, by estimating this LSTVAR model, we obtain two different sets of response parameters from two distinct regimes - a high expected inflation regime and a low expected inflation regime³⁶. The ESTVAR model, in contrast, offers a different economic interpretation to the LSTVAR model, where the two linear VARs in an ESTVAR model have similar dynamic structures in either of the two regimes. That is to say, if the expected inflation rates are different from the threshold value in the ESTVAR model, $G^i(.)$ will be different from zero, and the model will smoothly approximate from the middle ground (where the expected inflation equals the threshold value) to any of the two regimes, either has a high or low expected inflation rates.

V.8.2 Specification tests for the STVAR models

The specification tests for the STVAR models are similar to the univariate case described in section II.1.2 of Chapter II. Following Granger and Terasvirta (1993) and Camacho (2004), we first conduct the linearity tests against LSTVAR and ESTVAR (with the null being a linear VAR model), respectively. Second, we apply the model adequacy tests to the estimated model, which consists of tests for serially correlation in errors, tests for parameter constancy and tests for no remaining nonlinearity.

V.8.2.1 Linearity test

To apply the linearity test, we need to preselect a lag structure in equation (V.30) for both LSTVAR and ESTVAR models. This is done based on the lag structure of the linear VAR model, which suggests to include 4 lags. Next, we base both the linearity and model selection tests on Taylor series expansions of the transition function around $\gamma^i = 0$, and conducts the test statistics in the LM-type fashion. Table V-6 shows the linearization of the LSTVAR and ESTVAR models assuming two types of transition variables: (1) transition variables that are endogenous, which belongs to $\{y_{t-4}, x_{t-4}\}$ and (2) transition variables that are exogenous. To overcome the identification problem aforementioned in section II.1.2.1 of Chapter II and to discriminate the two models in respect of the model selection tests in the endogenous transition variable case, we use a second-order approximation for the ESTVAR model and a third-order

³⁶ Following the same logic, if the transition variable is defined as the lagged values of the stock returns, then the corresponding two regimes are high stock return regime (market expansion phrase) and the low stock return regime (market contraction phrase).

approximation for the LSTVAR model. For the exogenous transition variable case, we discriminate between the two models by applying a second-order Taylor expansion on the ESTVAR model and a first-order Taylor expansion on the LSTVAR model. The LM-type tests are then carried out on the auxiliary regressions depicted in the first column of Table V-6, with respect to the null of a linear VAR model when $\beta_1^i = \beta_2^i = \beta_3^i = 0$, $i = y, x$.

Table V-6: Linear approximation of LSTVAR and ESTVAR models

LSTVAR	ESTVAR
Transition variable is endogenous	
$y_t = u_t^y + \sum_{h=0}^3 \beta_h'^y X_t (s_t)^h + \varepsilon^y$	$y_t = u_t^y + \sum_{h=0}^2 \beta_h'^y X_t (s_t)^h + \varepsilon^y$
$x_t = u_t^x + \sum_{h=0}^3 \beta_h'^x X_t (s_t)^h + \varepsilon^x$	$x_t = u_t^x + \sum_{h=0}^2 \beta_h'^x X_t (s_t)^h + \varepsilon^x$
Transition variable is exogenous	
$y_t = \sum_{h=0}^1 \left(u_h^y (s_t)^h + \beta_h'^y X_t (s_t)^h \right) + \varepsilon^y$	$y_t = \sum_{h=0}^2 \left(u_h^y (s_t)^h + \beta_h'^y X_t (s_t)^h \right) + \varepsilon^y$
$x_t = \sum_{h=0}^1 \left(u_h^x (s_t)^h + \beta_h'^x X_t (s_t)^h \right) + \varepsilon^x$	$x_t = \sum_{h=0}^2 \left(u_h^x (s_t)^h + \beta_h'^x X_t (s_t)^h \right) + \varepsilon^x$

Note: $X_t = [\text{constant}, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}]$; ε^y and ε^x are the Taylor expansion remainders of the transition functions for y and x , respectively.

If the test of linearity rejects the null of a linear VAR model, we need to decide between a LSTVAR model and a ESTVAR model relying on a battery of model selection tests. In line with the univariate case of Granger and Terasvirta (1993) and following Camacho (2004), we show in Table V-7 the sequence of nested model selection hypothesis tests and the decision rules.

Table V-7: Nested model selection hypothesis tests

Transition variable is endogenous		
Test 1	H0: $\beta_3^i = 0$	H1: $\beta_3^i \neq 0$
Test 2	H0: $\beta_2^i = \beta_3^i = 0$	H1: $\beta_2^i \neq 0$, $\beta_3^i = 0$
Test 3	H0: $\beta_1^i = \beta_2^i = \beta_3^i = 0$	H1: $\beta_1^i \neq 0$, $\beta_2^i = \beta_3^i = 0$

Test 1	Reject H0	Accept H1	Accept H1	Accept H1
Test 2	...	Accept H1	Reject H0	Reject H0
Test 3	...	Reject H0	Accept H1	Reject H0
Decision	LSTVAR	LSTVAR	ESTVAR	Other
Transition variable is exogenous				
Test 1	H0: $u_2^i = 0, \beta_2^i = 0$		H1: $u_2^i \neq 0, \beta_2^i \neq 0$	
Test 1	Reject H0	Accept H1		
Decision	ESTVAR	LSTVAR		

V.8.2.2 Model adequacy tests

The model adequacy tests proposed by Eitrheim and Terasvirta (1996b) for univariate case and extended by Camacho (2004) to a multivariate framework consists of three tests - tests of serially correlated errors, tests for parameter constancy and tests for no remaining nonlinearity.

V.8.2.2.1 Test of serially correlated errors

To derive the LM test statistic for serially correlated errors, we reparameterize equation (V.30) as

$$\begin{aligned}
Y_t &= F(z_t, \Phi) + U_t, \\
Y_t &= [y_t, x_t]', \\
F(z_t, \Phi) &= [F^y(z_t, \Phi^y), F^x(z_t, \Phi^x)]', \\
U_t &= [u_t^y, u_t^x]',
\end{aligned} \tag{V.33}$$

where $F^i(z_t, \Phi^i) = z_t' \phi_1^i + (z_t' \phi_2^i) G(\gamma^i, c^i; s_t)$, and $\Phi^i = [\phi_1^i, \phi_2^i, \gamma^i, c^i]'$ is the vector of parameters to be estimated with $i = y, x$. Since the test of serially correlated errors tests if the model's residuals contain any remaining autocorrelation after taking into account the smooth transition effect, we postulate U_t , under the alternative hypothesis that it has serial correlation in it, evolves as

$$\begin{aligned}
U_t &= \Gamma(L)U_t + V_t, \\
V_t &\sim N(0, \Sigma),
\end{aligned} \tag{V.34}$$

where V_t is the serially independent errors and $\Gamma(L) = \Gamma_1 L^1 + \dots + \Gamma_r L^r$ is a 2×2 matrix polynomial in the lag operator L , r indicates the number of lagged residuals we consider in

the LM test. Under the null hypothesis of serially independent errors (i.e. $H_0 : \Gamma_1 = \dots = \Gamma_r = 0$), we can compute the LM test statistic, as shown in Camacho (2004), as

$$LM = \frac{1}{T} \frac{M'_\Gamma M_\Gamma}{M_{\Gamma\Gamma} - \frac{M_{\Gamma\Phi} M'_{\Gamma\Phi}}{M_{\Phi\Phi}}}, \quad (V.35)$$

where M_Γ , $M_{\Gamma\Gamma}$, $M_{\Phi\Gamma}$ and $M_{\Phi\Phi}$ are approximated as

$$\begin{aligned} M_\Gamma &= \sum (\underline{\Sigma}^{-1} \underline{U}_t \otimes \tilde{\underline{U}}_t) & M_{\Gamma\Gamma} &= \frac{1}{T} \sum (\underline{\Sigma}_t^{-1} \otimes \tilde{\underline{U}}_t \tilde{\underline{U}}_t') \\ M_{\Gamma\Phi} &= \frac{1}{T} \sum (\underline{\Sigma}_t^{-1} \underline{D}_t \otimes \tilde{\underline{U}}_t) & M_{\Phi\Phi} &= \frac{1}{T} \sum (\underline{D}_t \underline{\Sigma}_t^{-1} \underline{D}_t'). \end{aligned} \quad (V.36)$$

Specifically, a bar below any parameter expression refers to its maximum likelihood estimate under the null hypothesis of no serial correlation in the residuals. The variable $\tilde{\underline{U}}_t$ is a $2r \times 1$ matrix $[\tilde{u}_t^y, \tilde{u}_t^x]'$, where $\tilde{u}_t^i = [u_{t-1}^i, \dots, u_{t-r}^i]'$, with $i = y, x$. The variable $\underline{D}_t = [d_t^y, d_t^x]$, where d_t^i is $\partial F^i / \partial \Phi^i = \partial F^i(z_t, \Phi^i) / \partial \Phi^i$, with $i = y, x$. The LM test of serially correlated errors follows a χ^2 limiting distribution with $4r$ degrees of freedom³⁷.

V.8.2.2.2 Tests for parameter constancy

Since the STVAR models are estimated assuming constant parameters, a test for parameter constancy is important for checking the adequacy of the model. The test for our STVAR models is obtained by assuming the transition function has constant parameters but allowing ϕ_1^i and ϕ_2^i to change over time, that is we consider $\phi_j^i(t) = \phi_j^i + \lambda_j^i H_i(t)$ ($i = y, x$), where

$$H^i(t) = \frac{1}{1 + \exp(-\gamma^i (b_0^i + b_1^i t + \dots + b_{k-1}^i t^{k-1} + t^k))} - \frac{1}{2}. \quad (V.37)$$

To test whether ϕ_1^i and ϕ_2^i are time-varying, we need to linearly approximate equation (V.37), which gives us two auxiliary regressions for y and x :

³⁷ For the derivation of this LM test, we refer to Camacho (2004) for details.

$$\begin{aligned}
y_t &= \theta'_{1,0} \alpha_t^y + \theta'_{1,1} \alpha_t^y t + \dots + \theta'_{1,k} \alpha_t^y t^k + \\
&\left(\theta'_{2,0} \alpha_t^y + \theta'_{2,1} \alpha_t^y t + \dots + \theta'_{2,k} \alpha_t^y t^k \right) G(\gamma^y, c^y; s_t^y) + v_t^y \\
x_t &= \theta'_{1,0} \alpha_t^x + \theta'_{1,1} \alpha_t^x t + \dots + \theta'_{1,k} \alpha_t^x t^k + \\
&\left(\theta'_{2,0} \alpha_t^x + \theta'_{2,1} \alpha_t^x t + \dots + \theta'_{2,k} \alpha_t^x t^k \right) G(\gamma^x, c^x; s_t^x) + v_t^x,
\end{aligned} \tag{V.38}$$

where the null hypothesis of constant parameters implies $\theta'_{j,1} = \dots = \theta'_{j,k} = 0$, with $i = y, x$ and $j = 1, 2$.

V.8.2.2.3 Test for no remaining nonlinearity

Nonlinear models could be misspecified in many ways, consider the following additive misspecification of the STVAR model

$$\begin{aligned}
y_t &= z_t' \phi_1^y + (z_t' \phi_2^y) G_1^y(\gamma_1^y, c_1^y; s_t^y) + (z_t' \phi_3^y) G_2^y(\gamma_2^y, c_2^y; s_t^y) + u_t^y, \\
x_t &= z_t' \phi_1^x + (z_t' \phi_2^x) G_1^x(\gamma_1^x, c_1^x; s_t^x) + (z_t' \phi_3^x) G_2^x(\gamma_2^x, c_2^x; s_t^x) + u_t^x, \\
z_t &= [1, y_{t-1}, y_{t-2}, \dots, y_{t-p}, x_{t-1}, x_{t-2}, \dots, x_{t-p}]',
\end{aligned} \tag{V.39}$$

where the additive nonlinear component $G_2^i(\cdot)$ with $i = y, x$ can be either a logistic or exponential transition function and $G_2^i(0, c_2^i, s_t^i) = 0$. The test of no remaining nonlinearity starts from estimating the model without the second nonlinear component, and the null hypothesis $H_0: \gamma_2^i = 0$ is tested against (V.39)³⁸. Similar to the problem we encounter in the linearity test, equations (V.39) are not identified under the null hypothesis. Following Eitrheim and Terasvirta (1996b) and Camacho (2004), we apply linear approximation to $G_2^i(\cdot)$, which leads the auxiliary regressions

$$\begin{aligned}
y_t &= \beta_1'^y X_t + \beta_2'^y G_1^y(\cdot) + \beta_3'^y (s_t) + \beta_4'^y (s_t)^2 + \beta_5'^y (s_t)^3 + u_t^y \\
x_t &= \beta_1'^x X_t + \beta_2'^x G_1^x(\cdot) + \beta_3'^x (s_t) + \beta_4'^x (s_t)^2 + \beta_5'^x (s_t)^3 + u_t^x.
\end{aligned} \tag{V.40}$$

The null hypothesis of no remaining nonlinearity is then tested by considering $H_0: \beta_3^i = \beta_4^i = \beta_5^i = 0$, with $i = y, x$.

V.8.3 Empirical result

The data we use comprises the monthly stock return risk premium (nominal return less 3-month risk-free interest rate) from the FTSE100 index (r), obtained from Yahoo Finance,

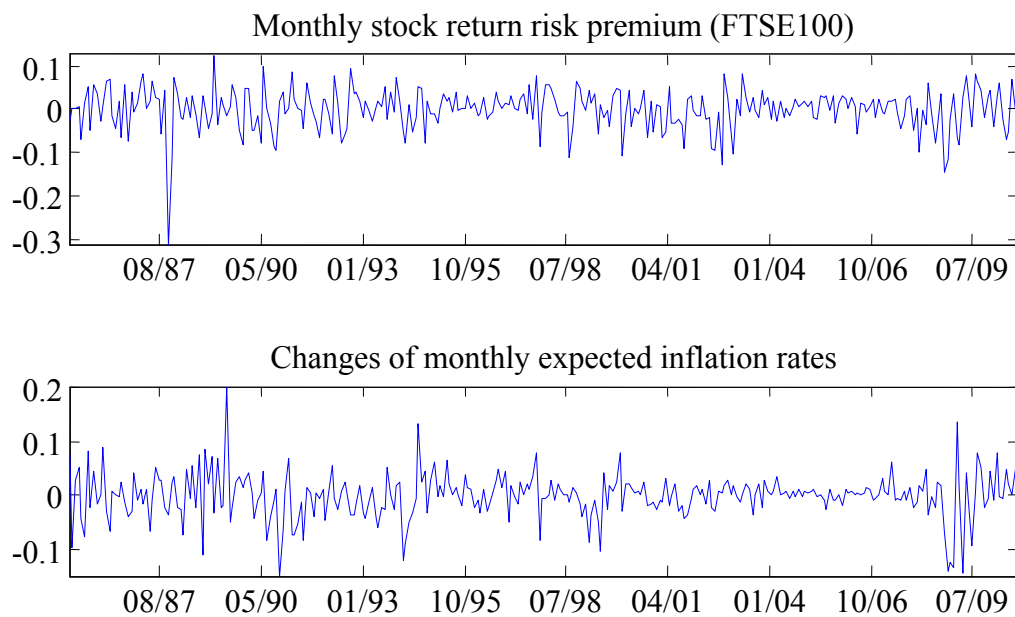
³⁸ We assume that under null hypothesis the parameters $\phi_j^i, \gamma_1^i, c_1^i$ can be consistently estimated.

and the monthly expected inflation rates obtained from section V.7. Since the preliminary data analysis has detected a unit root in the monthly expected inflation rates, we use the changes of the expected inflation rates ($\Delta\pi$) instead in the STVAR modelling. Both variables range from February 1985 to Oct 2010. The descriptive statistics and visual plots are shown in Table V-8 and Figure V-29, respectively.

Table V-8: Descriptive statistics of r and $\Delta\pi$

	r	$\Delta\pi$
Mean	-0.000743	-0.001618
Median	0.003462	0.000128
Maximum	0.125171	0.202029
Minimum	-0.309563	-0.150273
Std. Dev.	0.047157	0.042201
Skewness	-1.220625	-0.153744
Kurtosis	8.440945	6.288116
Normality (JB)	456.3994	139.9636
	[0.00000]	[0.00000]
Observations	308	308

Figure V-29: Monthly stock return risk premium (FTSE100) and changes of monthly expected inflation rates



Linearity test and model selection tests request a pre-specified linear VAR model, based on which we determine the correct number of lags of the dependent variables to be included into the STVAR models. The preliminary analysis suggests we include 4 lags of the dependent variables, which is selected by the AIC information criterion. In addition, we also need to pre-specify the transition variables in the LSTVAR and ESTVAR models. For this, we use r_{t-4} , r_{t-5} , $\Delta\pi_{t-4}$, $\Delta\pi_{t-5}$, $\hat{\pi}_{t-4}^u$ and $\hat{\pi}_{t-5}^u$ ³⁹, where $\hat{\pi}_{t-4}^u$ and $\hat{\pi}_{t-5}^u$ are the unexpected inflation rates obtained from Figure V-26 (unexpected inflation is calculated as the difference between the trend of the actual one month inflation rate and the expected inflation rate). The test statistics are shown in Table V-9. The first column describes the choice of transition variable in testing. Results in second column reveals that, for all transition variables considered, we reject linearity at 1% significance level. Columns 3 to 5 show us the model selection tests, and the last column gives the choices of models according to the selection rules depicted in Table V-7.

Table V-9: Results for linearity test and model selection test

	Linearity test	Test 1	Test 2	Test 3	Decision
$s_t = r_{t-4}$	91.10628 [0.00017]	20.65328 [0.19223]	8.74338 [0.92360]	61.70963 [2.7E-07]	LSTVAR
$s_t = \Delta\pi_{t-4}$	91.20839 [0.00017]	35.19158 [0.00374]	10.23569 [0.85403]	66.25250 [4.5E-08]	LSTVAR
$s_t = r_{t-5}$	74.37975 [0.00018]	19.28444 [0.37451]			LSTVAR
$s_t = \Delta\pi_{t-5}$	69.84547 [0.00061]	34.08015 [0.01231]			LSTVAR
$s_t = \hat{\pi}_{t-4}^u$	71.57412 [0.00038]	24.41952 [0.14176]			LSTVAR
$s_t = \hat{\pi}_{t-5}^u$	70.94631 [0.00045]	20.30372 [0.31598]			LSTVAR

Note: All tests are LM-type tests, which have χ^2 distributions with different degrees of freedom. p-values are reported in squared brackets.

In order to reduce the number of LSTVAR models, we select the models with strongest rejection of linearity within each transition variable family. This leads us to consider the models with $s_t = \Delta\pi_{t-4}$ and $s_t = r_{t-5}$. Table V-10 shows the maximum likelihood estimates of the model parameters, and Figure V-30 and Figure V-31 show the logistic transition functions

³⁹ However, one should not be limited to the chosen transition variables. We left it to our future work on testing other potential transition variables.

of the stock return risk premium equation from LSTVAR models with $s_t = \Delta\pi_{t-4}$ and $s_t = r_{t-5}$, respectively. The former model presents a transition speed parameter of 1.2184 that is larger than the one in the later model, which indicates a much sharper transition between regimes in the former model, as can be seen from the comparison of the two transition function plots. Although the 4-period lagged endogenous variables (both r_{t-4} and $\Delta\pi_{t-4}$) are significant in both r_t and $\Delta\pi_{t-4}$ equations, other lagged variables are generally not significant. The relations between the stock return risk premium and the lagged values of the changes in expected inflation are positive, albeit mostly are insignificant. The only one that has a significant forecasting effect on the return risk premium comes from $\Delta\pi_{t-4}$. In one regime, one unit increase in the expected inflation rate forecasts a negative impact on the return premium; while in the other regime, a same unit increase in the expected inflation rate forecasts a positive adjustment on the return premium. The estimated threshold that determines the sign of impact on return premium from changes of expected inflation rate is 1.3161. For example, if $\Delta\pi_{t-4}$ is higher than the threshold value 1.3161, the impact of expected inflation on return premium will be more likely to be a positive one. However, if $\Delta\pi_{t-4}$ falls below 1.3161, more likely the impact will be a negative one.

Table V-10: Maximum likelihood estimates of the LSTVAR models

LSTVAR model with $s_t = \Delta\pi_{t-4}$	
$\hat{r}_t =$	$\begin{pmatrix} -0.1065 + 0.1058 r_{t-1} - 0.3466 r_{t-2} + 0.0213 r_{t-3} - 0.0458 r_{t-4} \\ (0.0044) \quad (0.7316) \quad (0.4561) \quad (0.5034) \quad (0.0260) \\ + 0.0633 \Delta\pi_{t-1} + 0.1946 \Delta\pi_{t-2} + 0.0152 \Delta\pi_{t-3} - 0.3382 \Delta\pi_{t-4} \\ (0.1819) \quad (0.6773) \quad (0.4719) \quad (0.1763) \end{pmatrix}$ $+ \begin{pmatrix} 0.6279 - 0.2635 r_{t-1} + 1.4589 r_{t-2} - 0.3935 r_{t-3} + 0.8921 r_{t-4} \\ (0.0326) \quad (0.3030) \quad (1.6707) \quad (0.9583) \quad (0.5843) \\ + 0.5532 \Delta\pi_{t-1} - 1.9845 \Delta\pi_{t-2} + 0.3417 \Delta\pi_{t-3} + 0.2086 \Delta\pi_{t-4} \\ (1.1208) \quad (1.9803) \quad (0.9155) \quad (0.1638) \end{pmatrix} \hat{G}^r(.)$
$\Delta\hat{\pi}_t =$	$\begin{pmatrix} -0.4284 - 0.2718 r_{t-1} + 0.0457 r_{t-2} + 0.0599 r_{t-3} + 0.2592 r_{t-4} \\ (0.0384) \quad (0.1426) \quad (0.3975) \quad (0.5470) \quad (0.1266) \\ - 0.3302 \Delta\pi_{t-1} + 0.1670 \Delta\pi_{t-2} - 0.2294 \Delta\pi_{t-3} + 0.0749 \Delta\pi_{t-4} \\ (0.6155) \quad (0.3653) \quad (0.6489) \quad (0.0329) \end{pmatrix}$ $+ \begin{pmatrix} 0.8595 + 0.6496 r_{t-1} - 0.0463 r_{t-2} - 0.1710 r_{t-3} - 0.5977 r_{t-4} \\ (0.0768) \quad (1.3925) \quad (0.9016) \quad (1.1885) \quad (0.3170) \\ + 0.7361 \Delta\pi_{t-1} - 0.2503 \Delta\pi_{t-2} + 0.5143 \Delta\pi_{t-3} + 0.2013 \Delta\pi_{t-4} \\ (1.2845) \quad (0.7676) \quad (1.3502) \quad (0.1702) \end{pmatrix} \hat{G}^{\Delta\pi}(.)$
	$\hat{G}^r(.) = \left[1 + \exp\left(-\frac{1.2184}{0.3055}(\Delta\pi_{t-4} - \frac{1.3161}{0.3404})\right) \right]^{-1}$ $\hat{G}^{\Delta\pi}(.) = \left[1 + \exp(-0.3045(\Delta\pi_{t-4} + 0.0392)) \right]^{-1}$

$$\hat{\sigma}_{r_t r_t} = 0.0021, \hat{\sigma}_{\Delta\pi_t \Delta\pi_t} = 0.0017, \hat{\sigma}_{r_t \Delta\pi_t} = 0.0041$$

(1.5E-06) (1.8E-05) (0.0023)

LSTVAR model with $s_t = r_{t-5}$

$$\begin{aligned} \hat{r}_t &= \begin{pmatrix} 0.0193 - 0.1235 r_{t-1} - 0.2894 r_{t-2} - 0.1657 r_{t-3} + 0.0756 r_{t-4} \\ + 0.0735 \Delta\pi_{t-1} - 0.1184 \Delta\pi_{t-2} - 0.2657 \Delta\pi_{t-3} - 0.3864 \Delta\pi_{t-4} \end{pmatrix} \\ &+ \begin{pmatrix} -0.0545 + 0.5047 r_{t-1} + 0.6015 r_{t-2} + 0.2793 r_{t-3} + 0.1146 r_{t-4} \\ -0.2719 \Delta\pi_{t-1} + 0.7362 \Delta\pi_{t-2} + 0.7608 \Delta\pi_{t-3} + 0.5876 \Delta\pi_{t-4} \end{pmatrix} \hat{G}^r(.) \\ \Delta\hat{\pi}_t &= \begin{pmatrix} -0.3425 + 0.3482 r_{t-1} - 0.3192 r_{t-2} - 0.1659 r_{t-3} - 0.5447 r_{t-4} \\ -0.0431 \Delta\pi_{t-1} + 0.4942 \Delta\pi_{t-2} - 0.5076 \Delta\pi_{t-3} - 0.0345 \Delta\pi_{t-4} \end{pmatrix} \\ &+ \begin{pmatrix} 0.5114 - 0.3377 r_{t-1} + 0.4871 r_{t-2} + 0.2783 r_{t-3} + 0.7139 r_{t-4} \\ + 0.1243 \Delta\pi_{t-1} - 0.7255 \Delta\pi_{t-2} + 0.8201 \Delta\pi_{t-3} + 0.1678 \Delta\pi_{t-4} \end{pmatrix} \hat{G}^{\Delta\pi}(.) \\ \hat{G}^r(.) &= \left[1 + \exp \left(-0.5528 \left(r_{t-5} - 1.0252 \right) \right) \right]^{-1} \\ \hat{G}^{\Delta\pi}(.) &= \left[1 + \exp \left(-0.4967 \left(r_{t-5} + 1.4027 \right) \right) \right]^{-1} \\ \hat{\sigma}_{r_t r_t} &= 0.0021, \hat{\sigma}_{\Delta\pi_t \Delta\pi_t} = 0.0017, \hat{\sigma}_{r_t \Delta\pi_t} = 0.0041 \\ &\quad (5.2E-06) \quad (1.8E-05) \quad (0.0023) \end{aligned}$$

Note: standard errors are in parentheses

Figure V-30: Logistic transition function from LSTVAR model with $s_t = \Delta\pi_{t-4}$

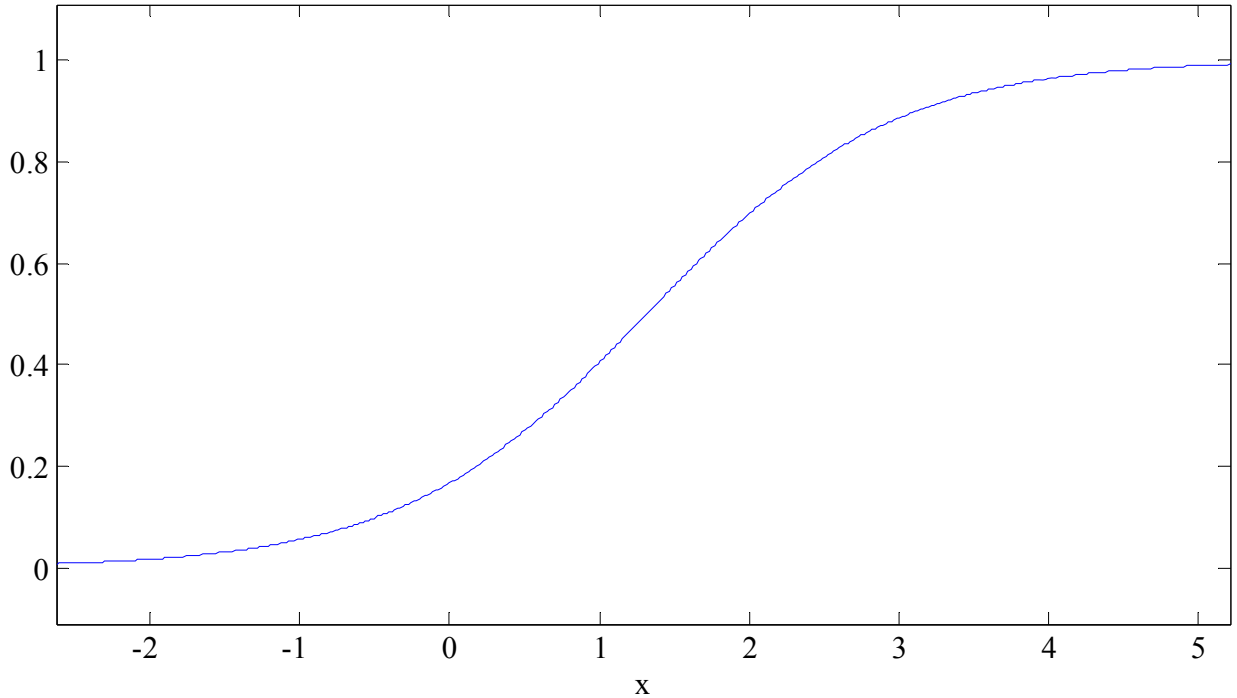
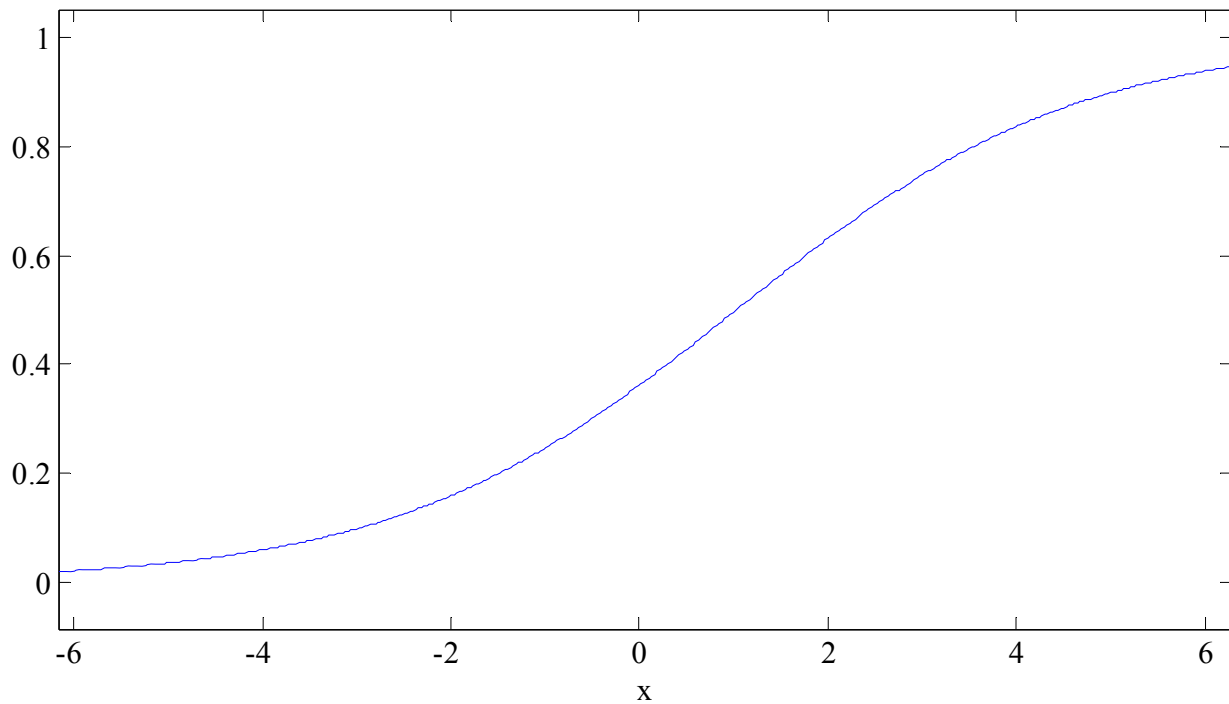


Figure V-31: Logistic transition function from LSTVAR model with $s_t = r_{t-5}$ 

In order to gauge the adequacy of the two nonlinear models in capturing the relation between stock return risk premium and the changes of the expected inflation rates, we carry out a series of model adequacy tests aforementioned. Table V-11 shows the results of the three model adequacy tests. Regarding to transition variable $s_t = \Delta\pi_{t-4}$, there still exists an autocorrelation problem in the errors, while the test of parameter constancy and test of no remaining nonlinearity show no evidence of parameter instability at $k=1$, and remaining nonlinearity. On the other hand, for the transition variable $s_t = r_{t-5}$, all three tests accept the null hypothesis of no serially correlated errors, parameter constancy and no remaining nonlinearity.

Table V-11: Model adequacy tests

	Test of serially correlated error test $r = 1$	Test of parameter constancy $k = 1$	Test of no remaining nonlinearity
$s_t = \Delta\pi_{t-4}$	28.16091 [1.2E-05]	11.6360 [0.99996]	43.02720 [0.67634]
$s_t = r_{t-5}$	7.38603 [0.11683]	28.86792 [0.79487]	14.60812 [0.99999]

Note: p-values are shown in square brackets.

To summarize the above findings on the relationship between stock return risk premium and changes of the expected inflation rate: 1) we learnt that postulating a linear relationship between the two variables is not convincing, as shown from the linearity tests,

there exists a nonlinear adjustment on the impact from lagged expected inflation rates to current return risk premium; 2) the consistent empirical rejection of the Fisherian hypothesis that assumes a positive relationship between stock market returns and inflation rates only captures a partial picture in a nonlinear regime-dependent setting; 3) in our STVAR framework, the switch from a negative relationship to a positive one depends on a threshold value of 1.3161; 4) if the transition variable $\Delta\pi_{t-4}$ is greater than this threshold value (although unlikely for low-inflation countries), the impact of lagged expected inflation rates on current return risk premium will be positive, hence the stock market hedges against the risks of rising inflation rates.

V.9 Conclusion

The first part of this study focuses on joint modelling of the nominal and real term structure of interest rates using both nominal bonds and inflation index-linked bonds. To decompose the nominal term structure of interest rates into real interest rates, expected inflation rates and inflation risk premia, we estimate a three-factor essentially affine no-arbitrage term structure model.

We find that the smooth expected month-on-month inflation rate rather behaves like the trend extracted from the observed series. On average, the expected series is above the trend before the year of 2000, but remains below it in the subsequent periods. Comparatively, the expected annual inflation rate also overstates the inflation rate before the new millennium and understates it in the following periods. This may suggest that the Bank of England's monetary policy is eventually quite creditable after it became independence in 1997. Most astonishingly, the recent sharp decline of the expected inflation rates (which even lead to a deflation) may lend support to the standing ground of the central bank to keep interest rates at historically low level.

In comparison with the historically low level of the expected inflation rates in recent periods, the 10-year inflation risk premia, however, is relatively high since the inception of the 2007/2008 financial crisis. By regressing the 10-year break-even rates on 3-month, 5-year and 8-year nominal yields for the periods January 1985 to December 2007 and January 2008 to September 2010, we find the adjusted R-square statistics in the second subsample is merely 39% compared to the 95% of the explanation power of the three nominal yields in the pre-crisis period, which suggests that there are other factors, other than the expected inflation and inflation risk premia, contained in the nominal yields during the financial crisis period. One

possible explanation on this is that the inflation risk premia estimated for this particular sample period may contain a large amount of liquidity premia.

In the second part of this study, we investigate the nonlinear relationship between stock return risk premium and the expected inflation rates, which are filtered out in the first part of this study. The nonlinearity test based on a STVAR framework shows that there exists a nonlinear adjustment on the impact from lagged inflation rates to current return premium. The only one that has a significant forecasting effect on the return risk premium comes from the 4-period lagged changes of expected inflation rate. In one regime, one unit increase in the expected inflation rate forecasts a negative impact on the return premium; while in the other regime, a same unit increase in the expected inflation rate forecasts a positive adjustment on the return premium. The changes of regimes, or the changes of signs on the impact of expected inflation rate to return premium is determined by the threshold value. While a larger change in the expected inflation rate forecasts a positive stock return premium, a modest change in the expected inflation rate predicts a negative response.

Bibliography

- Ait-Sahalia, Yacine, 1996, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385-426.
- Amihud, Yakov, 1996, Unexpected inflation and stock returns revisited--evidence from israel, *Journal of Money, Credit and Banking* 28, 22-33.
- Anderson, Nicola, and John Sleath, 2001, New estimates of the uk real and nominal yield curves, *Bank of England, Working paper No 126*.
- Ang, Andrew, Geert Bekaert, and Min Wei, 2007, Do macro variables, asset markets, or surveys forecast inflation better?*, *Journal of Monetary Economics* 54, 1163-1163.
- Ang, Andrew, Geert Bekaert, and Min Wei, 2008b, The term structure of real rates and expected inflation, *The Journal of Finance* 63, 797-797.
- Ang, Andrew, and Monika Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* 50, 745-787.
- Bank of England, 2009 Feb, Inflation report.

- Bansal, Ravi, and Hao Zhou, 2002, Term structure of interest rates with regime shifts, *Journal of Finance* 57, 1997-2043.
- Barclays Capital Research, 2008a. *Global inflation-linked products: A user's guide* (Barclays Capital).
- Barnes, Michelle L., 1999, Inflation and returns revisited: A tar approach, *Journal of Multinational Financial Management* 9, 233-245.
- Bernanke, Ben S., Vincent R. Reinhart, and Brian P. Sack, 2004, Monetary policy alternatives at the zero bound: An empirical assessment, *Brookings Papers on Economic Activity* 2004, 1-78.
- Benati, L., 2004, Evolving post-World War II U.K. Economic Performance, *Journal of Money, Credit, and Banking* 36, 691-718.
- Bodie, Zvi, 1976, Common stocks as a hedge against inflation, *Journal of Finance* 31, 459-470.
- Boudoukh, Jacob, and Matthew Richardson, 1993, Stock returns and inflation: A long-horizon perspective, *The American Economic Review* 83, 1346-1355.
- Boyd, John H., Ross Levine, and Bruce D. Smith, 2001, The impact of inflation on financial sector performance, *Journal of Monetary Economics* 47, 221-248.
- Buraschi, A., and A. Jiltsov, 2005, Inflation risk premia and the expectations hypothesis, *Journal of Financial Economics* 75, 429-490.
- Camacho, Maximo, 2004, Vector smooth transition regression models for us gdp and the composite index of leading indicators, *Journal of Forecasting* 23, 173+-173+.
- Campbell, John Y., and Robert J. Shiller, 1996, A scorecard for indexed government debt, *National Bureau of Economic Research Working Paper Series* No. 5587.
- Chen, Ren-Raw, and Louis Scott, 1993, Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* 3, 14-31.
- Chernov, Mikhail, and Philippe Mueller, 2008, The term structure of inflation expectations, *Bank of Canada, conference paper*.
- Christensen, J. H. E., J. A. Lopez, and G. D. Rudebusch, 2010, Inflation expectations and risk premiums in an arbitrage-free model of nominal and real bond yields, *Journal of Money Credit and Banking* 42, 143-178.
- Christopher, J. Neely, and E. Rapach David, 2008, Real interest rate persistence: Evidence and implications, (Federal Reserve Bank of St. Louis).
- Cochrane, John H., 2005, *Asset Pricing*, Revised Edition (Princeton University Press)

- Cochrane, John H., and Monika Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138-160.
- D' Amico, Stefania, Don H. Kim, and Min Wei, 2010, Tips from tips: The information content of treasury inflation-protected security prices, *Finance and Economics Discussion Series, Federal Reserve Board, Washington, D.C. Working Paper, 2010-19*.
- Deacon, Mark, Andrew Derry, and Dariusz Mirfendereski, 2004. *Inflation-indexed securities : Bonds, swaps and other derivatives* (John Wiley & Sons, Chichester).
- Duffee, Gregory R., 2002, Term premia and interest rate forecasts in affine models, *The Journal of Finance* 57, 405-443.
- Eitrheim, Øyvind, and Timo Terasvirta, 1996b, Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* 74, 59-75.
- Evans, Martin D. D., 1998, Real rates, expected inflation, and inflation risk premia, *Journal of Finance* 53, 187-218.
- Evans, Martin D. D., 2003, Real risk, inflation risk, and the term structure, *Economic Journal* 113, 345-389.
- Fama, E. F., and G. W. Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115-146.
- Fama, Eugene F., 1975, Short-term interest rates as predictors of inflation, *The American Economic Review* 65, 269-269.
- Fama, Eugene F., 1981, Stock returns, real activity, inflation, and money, *The American Economic Review* 71, 545-565.
- Fisher, Irving, 1930. *The theory of interest* (New York: The Macmillan Co.).
- García, Juan Angel , and Thomas Werner, 2010, Inflation risks and inflation risk premia, *ECB Working Paper Series, No 1162*.
- Garcia, Rene, and Pierre Perron, 1996, An analysis of the real interest rate under regime, *Review of Economics & Statistics* 78, 111-126.
- Granger, C. W. J., and Timo Terasvirta, 1993. *Modelling nonlinear economic relationships* (Oxford University Press, Oxford).
- Grishchenko, Olesya V., and Jing-Zhi Jay Huang, 2008, Inflation risk premium: Evidence from the tips market, (SSRN).
- Gultekin, N. Bulent, 1983, Stock market returns and inflation: Evidence from other countries, *Journal of Finance* 38, 49-65.

- Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright, 2010, The tips yield curve and inflation compensation.
- Hamilton, James D., 1985, Uncovering financial market expectations of inflation, *The Journal of Political Economy* 93, 1224-1242.
- Haubrich, J, G Pennacchi, and P Ritchken, 2008, Estimating real and nominal term structures using treasury yields, inflation, inflation forecasts, and inflation swap rates, *Federal Reserve Bank of Cleveland Working Paper No. 0810*.
- Hordahl, Peter, and Oreste Tristani, 2007, Inflation risk premia in the term structure of interest rates, *ECB Working Paper Series, No 734*.
- Jarrow, Robert, and Yildiray Yildirim, 2003, Pricing treasury inflation protected securities and related derivatives using an hjm model, *Journal of Financial and Quantitative Analysis* 38, 337-358.
- Joyce, Michael A. S., Peter Lildholdt, and Steffen Sorensen, 2010, Extracting inflation expectations and inflation risk premia from the term structure: A joint model of the uk nominal and real yield curves, *Journal of Banking & Finance* 34, 281-281.
- Kim, Don H., and Jonathan H. Wright, 2005a, An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates, *Finance and Economics Discussion Series, Federal Reserve Board, Washington, D.C. Working Paper, 2005-33*.
- Kim, J. R., 2003, The stock return-inflation puzzle and the asymmetric causality in stock returns, inflation and real activity, *Economics Letters* 80, 155-160.
- Lai, Kon S., 1997, Long-term persistence in the real interest rate: Some evidence of a fractional unit root, *International Journal of Finance & Economics* 2, 225-235.
- Lekkos, Ilias, and Costas Milas, 2004, Common risk factors in the u.s. And uk interest rate swap markets: Evidence from a nonlinear vector autoregression approach, *The Journal of Futures Markets* 24, 221-250.
- Lekkos, Ilias, Costas Milas, and Theodore Panagiotidis, 2007, Forecasting interest rate swap spreads using domestic and international risk factors: Evidence from linear and non-linear models, *Journal of Forecasting* 26, 601-601.
- Litterman, Robert, and Jose A. Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income*.
- Luintel, Kul B., and Krishna Paudyal, 2006, Are common stocks a hedge against inflation?, *The Journal of Financial Research* 29, 1-19.

- Piazzesi, Monika, 2010, Affine term structure models, in Yacine Ait-Sahalia, and Lars Hansen, eds.: *Handbook of financial econometrics*.
- Remolona, Eli M., Michael R. Wickens, and Frank F. Gong, 1998, What was the market's view of u.K. Monetary policy? Estimating inflation risk and expected inflation with indexed bonds, *FRB of New York Staff Report No. 57*
- Risa, Stefano, 2001, Nominal and inflation indexed yields: Separating expected inflation and inflation risk premia, (SSRN).
- Rose, Andrew K., 1988, Is the real interest rate stable?, *Journal of Finance* 43, 1095-1112.
- Rudebusch, G. D., B. P. Sack, and E. T. Swanson, 2007, Macroeconomic implications of changes in the term premium, *Federal Reserve Bank of St Louis Review* 89, 241-269.
- Rudebusch, Glenn D., and Tao Wu, 2008, A macro-finance model of the term structure, monetary policy and the economy, *The Economic Journal* 118, 906-926.
- Schmeling, Maik, and Andreas Schrimpf, 2010, Expected inflation, expected stock returns, and money illusion: What can we learn from survey expectations?, *European Economic Review* In Press, Corrected Proof.
- Sharpe, Steven A., 2002, Reexamining stock valuation and inflation: The implications of analysts' earnings forecasts, *Review of Economics and Statistics* 84, 632-648.
- Shen, Pu, 1998, How important is the inflation risk premium?, *Economic Review - Federal Reserve Bank of Kansas City* 83, 35-47.
- Summers, Peter M. 2005. What caused the Great Moderation? : some cross-country evidence, *Economic Review*, Federal Reserve Bank of Kansas City, issue Q III, pages 5-32.
- van Dijk, Dick, 1999, Smooth transition models: Extensions and outlier robust inference, Tinberg Institute (Amsterdam).
- Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177-188.

Appendix V

Table V-12: Review of recent estimations of inflation risk premia and inflation expectation

Author	Sample Periods	Country	Model	Data	IRP %	Inflation Expectation
ABW2008	1952-2004	US	Regime switching affine term structure model	Nominal yields and inflation rates	<ul style="list-style-type: none"> 5-year: 0.55% to 1.18% dependent on regimes unconditional 1.14% 	<ul style="list-style-type: none"> 5-year: 3.39% to 4.2% dependent on regimes unconditional 3.94%
Berardi2009	1960-2005	US	Macro-finance model	Nominal yields, real GDP and GDP deflator as proxy for price level	<ul style="list-style-type: none"> 1-year: 0.1% 10-year: 0.9% frequently go to negative till 1980 peaked during Volcker experiment period and then declined to around zero in 2006 	<ul style="list-style-type: none"> 1-year: 3.74% 5-year: 3.95%
BJ2005	1960-2000	US	Structural model	Nominal yields, money supply and inflation rate	<ul style="list-style-type: none"> 10-year: from 0.2% to 1.4%, 0.7% on average 5-year: 0.45% on average 	NA
CLC2010	1998-2007	US	2-factor CIR model	TIPS and nominal yields	<ul style="list-style-type: none"> 5-year: 0.2569% 10-year: 0.5561% 	annual inflation rate <ul style="list-style-type: none"> at 2.5%~ 3% in 2000-2001 and 2006-2007 around 0.5% in 2001-2005
CLR2010	2003-2008	US	4-factor No-arbitrage Nelson Siegel model	Weekly nominal yields and TIPS yields	<ul style="list-style-type: none"> 5 and 10 year: around 0 about 0.5% above/below 	<ul style="list-style-type: none"> 5-year: 1.93% to 2.54% 10-year: 2.15% to 2.40%
CM2008	1970-2004	US	No-arbitrage term structure model with observed macro variables	Nominal yields, surveys data, real per capita GDP and CPI	<ul style="list-style-type: none"> 1-year: 0.19% 10-year: 2.16% year rate fluctuates around zero 10-year rate fluctuates around 2.5% 	<ul style="list-style-type: none"> 1-year: 4.39% 10-year: 4.27% 1-year and 10-year peaked in 1980 and declined to 2.5% and around 2% respectively in 2005
CS1996	1953-1994	US	CAPM	Nominal yields, market portfolio and aggregate consumption	<ul style="list-style-type: none"> 5-year: from 0.7% to 1% 	NA
DKW2010	1990-2007	US	4-factor No-arbitrage affine term structure model	Weekly nominal yields and TIPS yields (since 2005), monthly inflation rates, monthly survey forecasts of short-term nominal interest rates and inflation	<ul style="list-style-type: none"> 10-year: from 0.5% to around 0 	<ul style="list-style-type: none"> 10-year: 3.5% to 2.5%
Evans1998	1983-1995	UK	No-arbitrage pricing kernel	Inflation index-linked gilt yields and	<ul style="list-style-type: none"> time-varying 	NA

				nominal yields, surveys of expected inflation rate	inflation risk premia	
GH2008	2000-2007	US	No-arbitrage pricing kernel	Monthly TIPS yields, CPI, standard real activity variables used in Macro VAR models, surveys of inflation rate	<ul style="list-style-type: none"> 10-year: from 0.11% to 0.22% negative rates in 2000-2003 positive rates in 2004-2007 	<ul style="list-style-type: none"> 5-year: around 2.5%
GW2010	1999-2006	Euro	No-arbitrage affine term structure model	Nominal and inflation index-linked gilt yields, inflation rates and survey data	<ul style="list-style-type: none"> 1-year: 0.07% 5-year: 0.25% 	<ul style="list-style-type: none"> 1-year: 1.94% 5-year: 1.9%
HPR2008	1982-2008	US	3-factor GARCH volatility central tendency term structure model	Nominal yields, inflation swap rates and survey forecasts of inflation	<ul style="list-style-type: none"> 10-year: varies from 0.38% to 0.6% 5-year: 0.27% 10-year: 0.51% 	NA
HT2010	US: 1990-2008 Euro: 1999-2008	US & Euro	Affine macro-finance model	Monthly nominal and real yields, inflation, output gap, and survey expectations of the short-term interest rate and inflation	<ul style="list-style-type: none"> US10-year: around 0, positive average rates before 2002 but zero afterwards Euro10-year: on average positive, highest (0.8%) in 2001-2002 to lowest in 2006- 2007, slightly increase recently but still around zero 	<ul style="list-style-type: none"> US10-year: from 3.5% declined to 2.5% Euro 10-year: from 1.875% decline until 2002 and then increased to around 2% in 2006 and then declined again
JLS2010	1992-2008	UK	No-arbitrage affine term structure model	Nominal and inflation index-linked gilt yields, RPI and Survey data	<ul style="list-style-type: none"> 5-year: around 1.25% (pre-1997), around 0.25% (post-1997) 10-year: around 1% (pre-1997), around 0.25% (post-1997) 	<ul style="list-style-type: none"> 5-year and 10-year pre-1997: around 3.5% post-1997: around 2.5%
Risa2001	1983-1999	UK	4-factor essentially affine term structure model	Weekly nominal and inflation index-linked gilt yields and RPI	<ul style="list-style-type: none"> 10-year: 2% initially at 4% and declines to around -1% 	<ul style="list-style-type: none"> 1-year: around 6% pre-1992; around 2% post-1992 20-year: declines from around 5% to around 3%
RWG1998	1982-1997	UK	2-factor affine term structure model	Nominal and inflation index-linked gilt yields	NA	<ul style="list-style-type: none"> 2-year: around 5% initially peaked at 9% in 1989-1990 and declined to 3% in 1995
Shen1998	1996-1997	UK	Modified Fisher equation	Surveys and target rate of monetary authorities	<ul style="list-style-type: none"> Survey calculation: 10-year: 0.74%; 25-year: 1.04%; Target calculation: 10-year: 1.33%; 25-year: 1.64% 	<ul style="list-style-type: none"> Survey calculation: 10-year: 3.106%; 25-year: 3.103 Target calculation: 10-year: 2.525%; 25-year: 2.51%

Matlab Code for joint modelling of the nominal and real term structure of interest rates

The following codes require CompEcon Toolbox to run. The toolbox can be downloaded from www4.ncsu.edu/~pfackler/compecon

```
%%=====run_uk.m=====
%% This is the main code file, it calls the likhfcn.m to minimize the
%% negative value of the likelihood value, calls the factor_filter.m to
%% filter out the factors and calculate risk premiums
%%=====
clear all
close all
warning off all;
global nominal_mat real_mat nominal_yield_no_error nominal_yield_with_error...
    real_yield_with_error nominal_ind_yields_no_error nominal_ind_yields_with_error...
    real_ind_yields_with_error total_t
%% LOAD DATA
% real yield from Jan 85 to Sep 10, maturities: 4y 5y 6y 7y 8y 9y 10y
% real rates 4y one point missing at Dec 1996

% nominal yield from Jan 85 to Sep 10, maturities: 3m 1y to 10y
% 3 month short rate from Jan 1985 to Feb 1997 is obtained from Datastream
% real and nominal rates are all from Bank of England
load UKReal4y_10yNom3m_10y_Jan85_Sep10.txt;
date_t=UKReal4y_10yNom3m_10y_Jan85_Sep10(:,1); % Jan 1985 to Sep 2010
data =UKReal4y_10yNom3m_10y_Jan85_Sep10(:,2:20);
total_t=size(data,1); % 1 to 309

% prepare real yields
real_yield=data(:,1:7)./1200;
real_mat=[48 60 72 84 96 108 120]; % 7 columns
nominal_yield=data(:,8:18)./1200;
nominal_mat=[3 12 24 36 48 60 72 84 96 108 120]; % 11 columns
obs_1m_inf=data(:,19);

%% CONTROL VARIABLES
% recursive=0;
number_of_factors=3;
number_of_unobserved_factors=3;

% select yields measured without errors and yields with errors
nominal_ind_yields_no_error=[3 5 10]; % 3 no error
nominal_ind_yields_with_error=[1 2 4 6 7 8 9 11]; % 8 with error
real_ind_yields_with_error=[1 2 3 4 5 6 7]; % 7 with error

nominal_yield_no_error=nominal_yield(:,nominal_ind_yields_no_error);
nominal_yield_with_error=nominal_yield(:,nominal_ind_yields_with_error);
real_yield_with_error=real_yield(:,real_ind_yields_with_error);

%% optimization

for initial_ind=1:1
    if initial_ind==1
        para_in=[0.95;-0.005;0.9;-0.002;-0.003;0.8;0.00001;0.00001;0.00001;0;0;0;0;0;0;0;0;0;0.003;0;0;1];
        options=optimset('Display','iter','TolX',1e-4,'TolFun',1e-8,...
            'MaxFunEvals',1000000000,'MaxIter',1000000000);
        [xout,fval,exitflag,output]=fminsearch(@likhfcn,para_in,options);
        %[xout,fval,exitflag,output]=fminunc(@likhfcn,para_in,options);
```



```

% estimation results
xfnl=para_trans(xout);
% S.E. and t-test
h_0=fdhess(@likhfcn,xout); % calculate Hessian matrix
g_0=fdjac1(@para_trans,xout,[]); % calculate gradient
h_fnl=g_0*(inv(h_0))*g_0'; % final Hessian
std_fnl=sqrt(diag(h_fnl)); % S.E.
t_ratio=xfnl./std_fnl; % t-ratio
% print results
table_full_sample_para=[xfnl std_fnl t_ratio];
sprintf('initial_ind=1')
sprintf('=====')
sprintf('Est.Para.~Std.Errors.~t-ratio')
disp(table_full_sample_para)
% save the first initial try
save('initial_ind=1');
bench_fval=-fval;
%=====

para_in=xout;
options=optimset('Display','iter','TolX',1e-4,'TolFun',1e-8,...
    'MaxFunEvals',1000000000,'MaxIter',1000000000);
[xout,fval,exitflag,output]=fminsearch(@likhfcn,para_in,options);
%[xout,fval,exitflag,output]=fminunc(@likhfcn,para_in,options);
% estimation results
xfnl=para_trans(xout);
% S.E. and t-test
h_0=fdhess(@likhfcn,xout); % calculate Hessian matrix
g_0=fdjac1(@para_trans,xout,[]); % calculate gradient
h_fnl=g_0*(inv(h_0))*g_0'; % final Hessian
std_fnl=sqrt(diag(h_fnl)); % S.E.
t_ratio=xfnl./std_fnl; % t-ratio
% print results
table_full_sample_para=[xfnl std_fnl t_ratio];
sprintf('initial_ind=2')
sprintf('=====')
sprintf('Est.Para.~Std.Errors.~t-ratio')
disp(table_full_sample_para)
% save the first initial try
save('initial_ind=2');
%=====

counter=3;
while -fval-bench_fval>=1
    bench_fval=-fval;
    para_in=xout;
    options=optimset('Display','iter','TolX',1e-4,'TolFun',1e-8,...
        'MaxFunEvals',1000000000,'MaxIter',1000000000);
    [xout,fval,exitflag,output]=fminsearch(@likhfcn,para_in,options);
    %[xout,fval,exitflag,output]=fminunc(@likhfcn,para_in,options);
    % estimation results
    xfnl=para_trans(xout);
    % S.E. and t-test
    h_0=fdhess(@likhfcn,xout); % calculate Hessian matrix
    g_0=fdjac1(@para_trans,xout,[]); % calculate gradient
    h_fnl=g_0*(inv(h_0))*g_0'; % final Hessian
    std_fnl=sqrt(diag(h_fnl)); % S.E.
    t_ratio=xfnl./std_fnl; % t-ratio
    % print results
    table_full_sample_para=[xfnl std_fnl t_ratio];
    counter_text=sprintf('initial_ind=%d',counter);
    sprintf('=====')

```

```

sprintf('Est.Para.~Std.Errors.~t-ratio')
disp(table_full_sample_para)
% save the first initial try
save(counter_text);
counter=counter+1;
end
else
para_in=para_in-0.01*randn(1,1); % randomly generate different initial values
options=optimset('Display','iter','TolX',1e-4,'TolFun',1e-8,...
'MaxFunEvals',1000000000,'MaxIter',1000000000);
[xout,fval,exitflag,output]=fminsearch(@likhfcn,para_in,options);
if -fval-bench_fval>=1
% estimation results
xfnl=para_trans(xout);
% S.E. and t-test
h_0=fdhess(@likhfcn,xout); % calculate Hessian matrix
g_0=fdjac1(@para_trans,xout,[]); % calculate gradient
h_fnl=g_0*(inv(h_0))*g_0'; % final Hessian
std_fnl=sqrt(diag(h_fnl)); % S.E.
t_ratio=xfnl/std_fnl; % t-ratio
% print results
table_full_sample_para=[xfnl std_fnl t_ratio];
try_label=sprintf('initial_ind=%d',initial_ind);
disp(try_label)
sprintf('=====')
sprintf('Est.Para.~Std.Errors.~t-ratio')
disp(table_full_sample_para)
save(try_label);
bench_fval=-fval;
end
end
end
%% FILTER OUT THE FACTORS
X=factor_filter(xout);
% calculate premia
[tp_nominal,tp_real,inf_risk_premia,exp_inf,y_nominal,y_real,inf_12m,real_short_rate,nominal_short_rate,inf_risk_premia_12m]=calculate_premia(xout,X);

function lik_val=likhfcn(para_in)
% This function calculates the real and nominal likelihood function
global nominal_mat real_mat nominal_yield_no_error nominal_yield_with_error...
real_yield_with_error nominal_ind_yields_no_error nominal_ind_yields_with_error...
real_ind_yields_with_error total_t

para=para_trans(para_in);

% state equation
phi1_1=para(1);
phi2_1=para(2);
phi2_2=para(3);
phi3_1=para(4);
phi3_2=para(5);
phi3_3=para(6);
phi=[phi1_1 0 0;
phi2_1 phi2_2 0;
phi3_1 phi3_2 phi3_3];

sig1=para(7);

```

```

sig2=para(8);
sig3=para(9);
sig=[sig1 0 0;
      0 sig2 0;
      0 0 sig3];

% market price of risk
lam0_1=para(10);
lam0_2=para(11);
lam0_3=para(12);
lam0=[lam0_1;lam0_2;lam0_3];

lam1_1=para(13);
lam1_2=para(14);
lam1_3=para(15);
lam2_1=para(16);
lam2_2=para(17);
lam2_3=para(18);
lam3_1=para(19);
lam3_2=para(20);
lam3_3=para(21);
lam1=[lam1_1 lam1_2 lam1_3;
      lam2_1 lam2_2 lam2_3;
      lam3_1 lam3_2 lam3_3];
delta0=mean(nominal_yield_with_error(:,1));
delta1_1=1;
delta1_2=1;
delta1_3=1;
delta1=[delta1_1;delta1_2;delta1_3];
deltapi0=para(22);
deltapi1_1=para(23);
deltapi1_2=para(24);
deltapi1_3=para(25);
deltapi1=[deltapi1_1;deltapi1_2;deltapi1_3];
%% measurement equation (bond pricing)
% nominal bond pricing, an is 1 by 120, bn is 3 by 120
[an,bn]=nominal_bond_pricing(delta0,delta1,sig,lam0,phi,lam1);
% real bond pricing, an_real is 1 by 120, bn_real is 3 by 120
[an_real,bn_real]=real_bond_pricing(delta0,delta1,sig,lam0,phi,lam1,deltapi0,deltapi1);

%% pick the bond prices for available yields
% an_select is 1 by 11
an_select=an(nominal_mat); % an_select => 3month ....10year parameter
% bn_select is 3 by 11
bn_select=bn(:,nominal_mat); % bn_select => 3month ....10year parameter
% an_real_select is 1 by 7
an_real_select=an_real(real_mat); % an_real_select => 4year....10year parameter
% bn_real_select is 3 by 7
bn_real_select=bn_real(:,real_mat); % bn_real_select => 4year....10year parameter

%% split the bond prices into two groups: with error and no error
% an_select_no_error is 1 by 3
an_select_no_error=an_select(nominal_ind_yields_no_error);
% bn_select_no_error is 3 by 3
bn_select_no_error=bn_select(:,nominal_ind_yields_no_error);
%ab_no_error=[an_select_no_error bn_select_no_error];

% an_select_with_error is 1 by 8
an_select_with_error=an_select(nominal_ind_yields_with_error);

```

```

% bn_select_with_error is 3 by 8
bn_select_with_error=bn_select(:,nominal_ind_yields_with_error);
%ab_with_error=[an_select_with_error bn_select_with_error];

% an_real_select_with_error is 1 by 7
an_real_select_with_error=an_real_select(real_ind_yields_with_error);
% bn_real_select_with_error is 3 by 7
bn_real_select_with_error=bn_real_select(:,real_ind_yields_with_error);
%ab_real_with_error=[an_real_select_with_error bn_real_select_with_error];

%% extract unobserved factors from yields measured without error
factors=zeros(total_t,3);
for j=1:total_t
    factors(j,:)=(nominal_yield_no_error(j,:)-an_select_no_error)/(bn_select_no_error);
end

%% the state dynamics
X=factors; % total_t by 3

%% the likelihood function
% -----no error part-----
% Item_1: -(T-1)*log(|det(J)|), det(J)=det(bn_select_no_error)
det_J=det(bn_select_no_error);
Item_1=-(total_t-1)*log(abs(det_J));

% -----unobserved state factor parts-----
% Item_2: -(T-1)/2 * log(det(sig2))
det_sig2=det(sig*(sig'));
Item_2=-((total_t-1)/2)*log(det_sig2);

% Item_3: -0.5*sum[(X_t-X_{t-1}*phi)'\*inv(sig2)*(X_t-X_{t-1}*phi)]
inv_sig2=(sig*(sig'))\eye(3); % 3 by 3
state_error_term=zeros(total_t-1,3);
state_error_term_square=zeros(total_t-1,3);
for jj=2:total_t
    state_error_term(jj,:)=X(jj,:)-X(jj-1,:)*phi';
    state_error_term_square(jj,:)=(state_error_term(jj,:).^2).*((diag(inv_sig2)))';
end
Item_3=-0.5*sum(sum(state_error_term_square),2);

% -----with error parts-----
% -----nominal
% Item_4: -0.5*(T-1)*log(sum(i:number_with_error,sig(i)^2))
nominal_yield_error_term=zeros(total_t-1,8);
for jjj=2:total_t
    nominal_yield_error_term(jjj,:)=nominal_yield_with_error(jjj,:)-...
        an_select_with_error-X(jjj,:)*bn_select_with_error; % 1 by 8
end
% 8 by 8 var-cov matrix
var_cov_nominal_yield_error_term=(nominal_yield_error_term'*nominal_yield_error_term)/(total_t-2);
var_nominal_yield_error_term=diag(diag(var_cov_nominal_yield_error_term));
Item_4=-0.5*(total_t-2)*sum(log(diag(var_nominal_yield_error_term)));

% Item_5
nominal_yield_error_term_square=nominal_yield_error_term.^2; % 120 by 8
nominal_weighted_square_residual=nominal_yield_error_term_square/var_nominal_yield_error_term; % 120
by 8
Item_5=-0.5*sum(sum(nominal_weighted_square_residual,2));

```

```

% -----real
% Item_6 same as Item_4
real_yield_error_term=zeros(total_t-1,7);
for jjjj=2:total_t
    real_yield_error_term(jjjj,:)=real_yield_with_error(jjjj,:)-...
        an_real_select_with_error-X(jjjj,:)*bn_real_select_with_error; % 1 by 7
end
% 7 by 7 var-cov matrix
var_cov_real_yield_error_term=(real_yield_error_term'*real_yield_error_term)/(total_t-2);
var_real_yield_error_term=diag(diag(var_cov_real_yield_error_term));
Item_6=-0.5*(total_t-2)*sum(log(diag(var_real_yield_error_term)));

% Item_7 same as Item_5
real_yield_error_term_square=real_yield_error_term.^2; % total_t by 8
real_weighted_square_residual=real_yield_error_term_square/var_real_yield_error_term; % 120 by 8
Item_7=-0.5*sum(sum(real_weighted_square_residual,2));

%=====sum the log likelihood terms=====
lik_val=-(Item_1+Item_2+Item_3+Item_4+Item_5+Item_6+Item_7);

% constraint violations
if abs(imag(lik_val))>0
    lik_val=real(lik_val)+1e8;
end
end % of function

function [an,bn]=nominal_bond_pricing(delta0,delta1,sig,lam0,phi,lam1)
% This function calculate the nominal bond prices
An=zeros(1,120);
Bn=zeros(3,120);

% when n=1, 1 period of maturity left
An(1)=-delta0;
Bn(:,1)=-delta1;
an=-An;
bn=-Bn;

for i=2:120
    An(i)=An(i-1)-Bn(:,i-1)*sig*lam0+0.5*Bn(:,i-1)*sig*(sig)*Bn(:,i-1)-delta0;
    Bn(:,i)=(Bn(:,i-1)*(phi-sig*lam1)-delta1)';
    an(i)=-An(i)/i;
    bn(:,i)=-Bn(:,i)/i;
end
end % of function

function [an_real,bn_real]=real_bond_pricing(delta0,delta1,sig,lam0,phi,lam1,deltapi0,deltapi1)
% This function calculates the real bond prices
An_real=zeros(1,120);
Bn_real=zeros(3,120);

% when n=1, 1 period of maturity left
An_real(1)=-(delta0-deltapi0);
Bn_real(:,1)=-(delta1-deltapi1);
an_real=-An_real;

```

```

bn_real=-Bn_real;

for i=2:120
    An_real(i)=An_real(i-1)-Bn_real(:,i-1)*sig*lam0+0.5*Bn_real(:,i-1)*sig*(sig)*Bn_real(:,i-1)-(delta0-
deltapi0);
    Bn_real(:,i)=(Bn_real(:,i-1)*(phi-sig*lam1)-(delta1-deltapi1))';
    an_real(i)=-An_real(i)/i;
    bn_real(:,i)=-Bn_real(:,i)/i;
end
end % of function

```

Chapter VI The stock market risk return Trade-off: evidence from six countries using centuries-long data

VI.1 Introduction

The risk-return trade-off is a central issue in financial economics. Are the expected returns of stock market investment predictable over time and, if so, is this predictability related to the stock market volatilities? How is the relation between the expected return of stock market investment and its volatility? Is this relation negative or positive, and is this relation stable over time? The progression of numerous empirical evidence attempted answering the above questions has been continuously challenging the asset pricing theory, which served as both academic researchers and market practitioners' interests.

In this study, we provide additional insight into the nature of the aggregate stock market volatilities and its relations to expected returns, in a Markov switching model using centuries-long aggregate stock market data from six countries (Australia, Canada, Sweden, Switzerland, the United Kingdom and the United States). As Lundblad (2007) argued, the primary challenge in estimating the risk-return relation is the small sample used in previous studies. In a Monte Carlo analysis, the author claims that one requires an extremely long history of the stock return data to detect a significant risk-return trade-off within the GARCH-in-mean framework. In comparison to his study on the US data solely, our dataset covers six countries, which all have century-long history, and hence gives us the advantage to take a broader view of the risk-return trade-off both horizontally across countries and vertically in an extremely long period.

Since there is no strict theoretical premise about the correct model for stock returns, we consider a wide range of models in this study. By relaxing restrictions on the multi-regime switching parameters in each step, we estimated a series of models, either allowing each of the expected mean and volatility of the stock return to switch or permitting both moments of the stock return to shift regimes over time. In contrast to most studies in Markov switching modelling of stock returns that assuming two regimes, we also consider three regimes for the variation of stock returns. This seemingly subtle complication, as compared to the conventional two regimes assumption, allows a richer inter-temporal relation between the expected mean and volatility of the stock returns. We find that the Markov switching models assuming both regime dependent mean and volatility with a 3-regime specification are more

capable to captures the extreme movements of the stock market which are short-lived. This helps us to capture the stock return's leptokurtosis in its distribution. In addition, as most of the extreme movements in the stock market are downwards, Markov switching models with 3-regime specification would better capture the negative skewness in the stock returns than their 2-regime counterparts.

We also investigate into the "volatility feedback effect" hypothesis that has been documented extensively in literature. Following Turner, Startz and Nelson (1989) and Kim, Morley and Nelson (2004), we make different assumptions about the information available to market investors. By learning about the current prevailing volatility regime based on different information sets prior to the current trading period, we are able to see the effect on the additional risk premium required by investors in anticipating future volatility changes. We find the volatility feedback effect that we studied on these six countries shows a positive sign on anticipating a high volatility regime of the current trading month. In other words, having taken account for the volatility feedback effect, the empirical evidence shows that stock prices move in the opposite direction to the level of market volatility which creates an immediate realized negative stock return.

The remainder of the study is organized as follows. Section 2 provides a literature review of the risk-return trade-off in stock market. In Section 3, the five econometric models estimated in this study are presented. Section 4 provides the description of the data series used. Section 5 gives the estimation results of the econometric models for each country. Finally, Section 6 concludes.

VI.2 Literature review

The classical asset pricing models, such as the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965), explicitly claim that periods of high stock returns should coincide with periods of high market volatility, implying a positive correlation between the stock market returns and their volatility. However, the body of empirical evidence on the stock market risk-return relationship is mixed and inconclusive. Some studies (e.g. French, Schwert and Stambaugh (1987), Bollerslev, Engle and Wooldridge (1988), Harvey (1989b), Harrison and Zhang (1999), Bansal and Lundblad (2002), Ghysels, Santa-Clara and Valkanov (2005), Ludvigson and Ng (2007) and Lundblad (2007)) find positive risk-return trade-off, but others (e.g. Campbell (1987), Breen, Glosten and Jagannathan (1989), Glosten,

Jagannathan and Runkle (1993), Whitelaw (1994) and Brandt and Kang (2004)) suggests negative relationships.

The empirical findings of negative risk-return trade-off, which is contrary to the prediction of mainstream of financial theory, has two competing economic explanations - the "leverage effect" hypothesis and the "volatility feedback effect" hypothesis. The "leverage effect" hypothesis, which can be traced back to Black (1976), states that the drop of the market price of a stock will cause the leverage of the firm to increase, and as a consequence, the volatility of the stock price will also increase. Although this hypothesis is able to explain the negative relationship between the realized stock returns and volatility, subsequent studies such as Christie (1982), Schwert (1989), Campbell and Hentschel (1992) and Bekaert and Wu (2000) argue that the leverage effect alone is too small to account for the negative risk-return trade-off, while the "volatility feedback effect" appears more plausible. The "volatility feedback effect" hypothesis essentially arises from the inter-temporal asset pricing models traced back to Merton (1973). It states that an increase in conditional stock market volatility would lead to a further increase in volatility if the volatility is persistent over time (which has been documented extensively in the literature, see e.g. Bollerslev, Chou and Kroner (1992)). Since this anticipated increase of future market volatility requires a higher market return in the future, the stock price should immediately adjust in the opposite direction to the level of the market volatility. Many studies on the "volatility feedback effect" (e.g. French, *et al.* (1987), Turner, *et al.* (1989), Campbell and Hentschel (1992) and Kim, *et al.* (2004)) suggest that taking volatility feedback effect into account is important to avoid confusing a negative risk-return trade-off with the true underlying relationship between realized return and the market volatility. Furthermore, volatility feedback can explain why stock market returns tends to be negatively skewed due to the asymmetric effect on positive and negative stock return news. Specifically, Campbell and Hentschel (1992) postulates that volatility increasing, given a positive stock return shock, feedbacks a decrease in prices which dampens the initial positive stock return shock. An initial negative stock return shock, however, would be amplified through the additional decreases in stock prices from the volatility feedback effect. As a result, stock returns are negatively skewed.

Empirical modelling of the risk-return trade-off usually relates the conditional mean of stock returns to conditional volatility in a linear regression fashion (e.g. as the risk-return trade-off representation in the CAPM). A popular approach for modelling the conditional mean is to use projections of stock returns onto predetermined conditioning variables and

taking the fitted values as the conditional mean. The conditional variance can be obtained by collecting the fitted values from a projection of the squared residuals (which are collected from the previous regression) onto the same set of predetermined conditioning variables. This approach is adopted by Campbell (1987) and Breen, *et al.* (1989). Alternatively, studies of Schwert (1989), Whitelaw (1994), Bollerslev, Litvinova and Tauchen (2006) and Ludvigson and Ng (2007) estimate conditional variance of stock returns using realized high-frequency daily return data (or intraday volatility data). Yet, the most popular approach to estimate conditional variance of the stock return is by specifying a parametric model, such as the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model of Engle (1982) and Bollerslev (1986). Although the GARCH model is able to capture the persistent volatility of stock returns as demonstrated in Monte Carlo simulations in Chou (1988), it assumes a constant mean in stock return which is incapable to capture the volatility feedback effect. To circumvent this drawback of the GARCH model, the GARCH-in-Mean (GARCH-M) model of Engle, Lilien and Robins (1987) allows the conditional mean of the stock returns to dependent on the conditional variance. Studies using GARCH-M model to estimate the relation between conditional mean and variance of stock returns usually find insignificant relation in international stock markets. Although French, *et al.* (1987) documented a significant positive relation in the US stock market, Baillie and DeGennaro (1990) argue that such a positive relation is weak. The estimated insignificant relation between the conditional mean and variance of stock return, as many argued, is due to the lack of a proper measure of conditional variance. The GARCH-M model, however, is sensitive to model misspecification, which would contaminate the estimated conditional variance. As argued in Nelson (1991), parameter restrictions imposed by GARCH models may restrict the dynamics of the conditional variance process. Other extensions of the GARCH model class, e.g. the Asymmetric GARCH (AGARCH) model of Engle and Ng (1993), the Asymmetric GARCH-M (AGARCH-M) model of Campbell and Hentschel (1992), the GJR-GARCH model of Glosten, *et al.* (1993) and the exponential-GARCH (EGARCH) model of Nelson (1991), may also have the same difficulties in obtaining a fully correctly specified model of the conditional variance. Indeed, a recent study by Lundblad (2007) argues that " ... the choice of volatility specification in the GARCH-M context is of second-order importance given the extremely low R^2 ...conditional volatility has almost no explanatory power for future realized returns". The primary challenge in estimating the risk-return relation, as the author suggested, is not related to the volatility specification, but the small sample used in various studies.

In contrast to the voluminous risk-return trade-off studies based on GARCH class models, relatively little attention has been paid to the capability of Markov switching class of models in exploring the risk-return relation in stock market returns. Early works by Turner, *et al.* (1989) examined a Markov switching model of stock return and variance, in which the market investors are uncertain about the current prevailing state of the variance regime. They find market investors are consistently surprised by high-variance regimes which implies a positive correlation between the risk premium and anticipated future volatility regimes. Kim, *et al.* (2004) extended to develop a formal "volatility feedback effect" model under Markov switching market volatility by incorporating a log-linear present value framework of Campbell and Shiller (1988). Their empirical results suggest a negative and significant "volatility feedback effect", which they argue to support a positive relation between stock market volatility and the return. A more recent study by Bae, Kim and Nelson (2007) supports this conclusion. By controlling for leverage as in Bekaert and Wu (2000), they find the changes in volatility regime account for most of the persistence in volatility process, indicating that the recurrent regime shifts are the main source of the negative correlation between realized returns and subsequent volatilities, and the intra-regime volatility feedback effect, which are captured by GARCH, weakens once the regime shifts are incorporated. Additionally, their findings are consistent with Mayfield (2004) which suggests that market investors also require risk premium associated with the risk of future changes in volatility regime in addition to the risk premium associated with conditional volatility.

VI.3 Econometric models

Let's denote each country's stock return series as r_{it} , where i is the country index, $i \in \{AUS, CAN, SWE, SWI, UK, US\}$. Assuming r_{it} follows a Markov switching process where the parameters are driving by an unobserved state variable, S_{it} , which is unique to each asset returns. S_{it} takes values $S_{it} = 1, \dots, k_i$, where k_i is the number of states for the i^{th} return series. We consider different econometric specifications on the DGP of each univariate return series. Each of the later models is building on earlier models by relaxing restrictions on parameter spaces.

VI.3.1 Single regime mean and variance (Model 1)

$$r_{it} - \mu_i = \sum_{j=1}^p \alpha_{ij} (r_{i,t-j} - \mu_i) + \sigma_i \varepsilon_{it}, \quad \varepsilon_{it} \sim iid N(0,1) \quad (VI.1)$$

The benchmark model, which has a constant mean (μ_i) and variance (σ_i) under a single regime, describes each demeaned univariate return series in terms of white noise and past periods' returns. If $\alpha_{ij} = 0$, this model implies that stock returns are homoskedastic and normal. If $\alpha_{ij} \neq 0$, this model implies that past returns have some predictability on future returns, which is similar to Jegadeesh (1991)'s mean reversion regression. The basic Jegadeesh (1991) model, which takes the form of $r_t - \mu = \alpha(p) \sum_{j=1}^p (r_{t-j} - \mu) + \varepsilon_t$, explains the difference between the one-period return r_t and its unconditional mean μ by the lagged returns r_{t-j} , where p is the holding period for lagged returns. If the coefficients, $\alpha(p)$, of these lagged returns equal zero then the returns are serially uncorrelated with expected value μ . Alternatively, if $\alpha(p) < 0$, it suggests the presence of mean reversion in stock prices. The difference, however, between (VI.1) and Jegadeesh (1991)'s mean reversion regression is that the holding period of past returns in his specification is greater than one period and hence is conventionally known as long horizon predictability regression (i.e. a sequence of accumulated one-period past returns over a holding period of p forecasts current one-period return).

VI.3.2 Single regime mean and multi-regime variance (Model 2)

$$\begin{aligned}
 r_{it} - \mu_i &= \sum_{j=1}^p \alpha_{ij} (r_{i,t-j} - \mu_i) + \sigma_{S_{it}} \varepsilon_{it}, \quad \varepsilon_{it} \sim iid N(0,1) \\
 \sigma_{S_{it}} &= \sum_{k=1}^m S_{ikt} \sigma_k \\
 S_{ikt} &= 1 \text{ if } S_{it} = k; \quad S_{ikt} = 0 \text{ if } S_{it} \neq k \\
 \Pr[S_{it} = k | S_{it-1} = k] &= p_{kk}
 \end{aligned} \tag{VI.2}$$

By relaxing the restriction of constant variance in Model 1, we allow the return series in model 2 to have multiple variance regimes depending on the unobserved state variable process (S_{it}) that characterizing different economic conditions in market. This state variable is a first-order, homogeneous, irreducible and ergodic Markov chain. The transition between regimes is governed by the transition probability, which is assumed to be constant. For each univariate return series, the number of state variables (and hence the number of regimes), k , is an important parameter for that it allows us to fit more complicated return dynamics by simply increasing it to 2 or 3. If $k = 1$, we are back to single regime model - model 1.

To defend the multi-regime variance specification, we note that French, *et al.* (1987), Schwert (1989) and many others have documented that large changes in variance of stock returns (in other words, the heterogeneity) may be well captured by regime changes in variance. In other words, the observed peaks and heavy tails of stock returns, as mentioned in Turner, *et al.* (1989), are typical of unconditional densities of normal observations subject to heteroskedasticity, where some of the sample observations are drawn from normal densities with smaller variances than the sample variance and some observations are drawn from normal densities with larger variances than the sample variance. Since the unconditional density of the sample observations is a linear combination of the normal densities with smaller and larger variances, it has more mass in its tails and peak than a simple normal distribution. On the other hand, volatility models based on autoregressive conditional heteroskedasticity (ARCH) have been commonly used in capturing some of the stylised facts of financial market variables. The GARCH model developed by Engle (1982) and Bollerslev (1986) are specifically designed to capture volatility clustering of returns. Although the GARCH effects may be highly significant with daily and weekly data, as suggested by Bollerslev, Chou and Kroner (1992), its effects tends to be much milder in less frequently sampled data. Using weekly data, Hamilton and Susmel (1994) also shows that the persistent ARCH effect completely dies out after a month. They also suggest the long-run unconditional variance is subjected to regime switches, which may give us sufficient reason to model the monthly sampled stock returns in a pure Markov switching variance model.

VI.3.3 Multi-regime mean and variance (Model 3)

$$\begin{aligned}
 r_{it} - \mu_{S_{it}} &= \sum_{j=1}^p \alpha_{ij} (r_{i,t-j} - \mu_{S_{i,t-j}}) + \sigma_{S_{it}} \varepsilon_{it}, \quad \varepsilon_{it} \sim iid N(0,1) \\
 \sigma_{S_{it}} &= \sum_{k=1}^m S_{ikt} \sigma_k, \quad \mu_{S_{it}} = \sum_{k=1}^m S_{ikt} \mu_k \\
 S_{ikt} &= 1 \text{ if } S_{it} = k; \quad S_{ikt} = 0 \text{ if } S_{it} \neq k \\
 \Pr[S_{it} = k | S_{it-1} = k] &= p_{kk}
 \end{aligned} \tag{VI.3}$$

In model 3, we consider regime switches in expected returns, stemming from changes in market risk premium as return variances switches between regimes. Particularly, we link the mean of the stock return to the variance regimes. Although this implies that the recurrent regime switches affect the systematic risk component of a well diversified portfolio such as those we study here, a large literature suggest the stock return risk premium is time dependent.

VI.3.4 Multi-regime mean, variance and autoregressive coefficient (Model 4)

$$\begin{aligned}
 r_{it} - \mu_{S_{it}} &= \sum_{j=1}^p \alpha_{S_{it},j} \left(r_{i,t-j} - \mu_{S_{i,t-j}} \right) + \sigma_{S_{it}} \varepsilon_{it}, \quad \varepsilon_{it} \sim iid N(0,1) \\
 \sigma_{S_{it}} &= \sum_{k=1}^m S_{ikt} \sigma_k, \quad \mu_{S_{it}} = \sum_{k=1}^m S_{ikt} \mu_k, \quad \alpha_{S_{it},j} = \sum_{k=1}^m S_{ikt} \alpha_{k,j} \\
 S_{ikt} &= 1 \text{ if } S_{it} = k; \quad S_{ikt} = 0 \text{ if } S_{it} \neq k \\
 \Pr[S_{it} = k \mid S_{it-1} = k] &= p_{kk}
 \end{aligned} \tag{VI.4}$$

The difference between model 4 and model 3 is the addition of a regime dependent autoregressive coefficient. The presence of this autoregressive term in previous models is used to proxy for omitted variables that tracts time-dependent stock return risk premium. Although we included this autoregressive coefficient as a constant due to our ambiguity about the correct theoretical model, it is interesting to see if the regime dependent risk premium in each regime is still predictable. As mentioned in Turner, *et al.* (1989), a properly specified model of risk premium must allow a time-dependent variance with a predictable element, which implies that the risk premium is time-dependent. If the time-dependent risk premium properly taken into account by the regime switching mean in model 3, we would expect insignificant regime-dependent α s in (VI.4).

VI.3.5 Multi-regime mean, variance and autoregressive coefficient with learning (Model 5)

$$\mu_{S_{it}} = (1 - S_{i2t}) \mu_1 + S_{i2t} \mu_2 + \theta \Pr[S_{i2t} = 1 \mid \Psi_{t-1}] \tag{VI.5}$$

The final model that we consider involves the learning process of market investors. Specifically, we make different assumptions about market investors' information set in model 5 comparing to those in previous models. In models 2 to 4, we assume the market investors know exactly what the current regime is, even though the econometricians do not observe it. In model 5, however, we assume neither market investors nor the econometricians observe the true current regime of the stock market, but have to form the probabilities of experiencing each possible regimes using information in the past. Turner, *et al.* (1989) in their paper emphasised that allowing market investors to learn the current regime of the market that it is staying on is necessary to reveal the empirical evidence of a positive risk-return trade off. The financial logic behind this says that a positive expected stock return risk premium is usually associated with persistent anticipated volatility increases as elementary financial theory suggests. If, however, there is an unanticipated stock return volatility hike, market investors would bid prices downwards which will make future expected returns higher enough to

compensate for the increase in this unanticipated non-diversifiable risk. This unanticipated systematic risk hike then will lead to a one-time negative return as market investors learn about the current regime of the stock return volatility. Therefore, in order to capture the true risk premium, a properly structured model needs to separate the effect of a negative return associated with the learning of market investors about the unanticipated volatility increases from the effect of a positive return associated with persistent and fully anticipated volatility increases.

To incorporate the learning process of market investors about the unanticipated stock return volatility, we follow Turner, *et al.* (1989) to model the prior and posterior probabilities of the Markov chain process that determines the volatility regimes. Specifically, we assume the market investors base their trading decisions at the beginning of the trading period t on the prior distribution of the state in that period, which are obtained using information up to $t-1$. Through the trading period, investors update their beliefs about the prevailing regime of states (posterior distribution of the states) based on the information set at time t using Bayes' theorem. Denoting the information set at the beginning of the trading period t as Ψ_{t-1} , the market investors' prior distribution of the state is $\Pr[S_{it} = k | \Psi_{t-1}] = \Pr[S_{ikt} = j | \Psi_{t-1}]$, $j = 0, 1$. In the case of two volatility regimes that we consider here (i.e. $k = 2$), the prior distribution can be summarized by the prior probability of a high volatility state⁴⁰ that is $\Pr[S_{i2t} = 1 | \Psi_{t-1}]$, which is equal to the expected value of the state variable conditioned on information set Ψ_{t-1} . During the trading period t , investors form the posterior distribution of the states by updating their prior distribution using

$$\Pr[S_{i2t} = 1 | \Psi_t] = \frac{\Pr[S_{i2t} = 1 | \Psi_{t-1}] \times f(\Psi_t | S_{i2t} = 1, \Psi_{t-1})}{f(\Psi_t | \Psi_{t-1})},$$

where $f(\Psi_t | S_{i2t} = 1, \Psi_{t-1})$ is the density of the information set conditional on the state of the system being in a high volatility state, and $f(\Psi_t | \Psi_{t-1})$ is the unconditional density of the information set. The Markov chain property of the state process allow us to calculate the prior distribution of the state in the following period as a linear transformation of the posterior $\Pr[S_{i2t} = 1 | \Psi_t]$ as

⁴⁰ A high volatility state is represented by $S_{i2t} = 1$.

$$\begin{aligned}\Pr[S_{i2t+1} = 1 | \Psi_t] &= \Pr[S_{i2t+1} = 1 | S_{i2t} = 0] \times \Pr[S_{i2t} = 0 | \Psi_t] \\ &+ \Pr[S_{i2t+1} = 1 | S_{i2t} = 1] \times \Pr[S_{i2t} = 1 | \Psi_t],\end{aligned}$$

where $\Pr[S_{i2t+1} = 1 | S_{i2t} = 0]$ and $\Pr[S_{i2t+1} = 1 | S_{i2t} = 1]$ are the transition probabilities of the states switching to a high volatility regime in trading period $t + 1$ conditional on the states are in low and high volatility regime, respectively, in the previous trading period.

Given the learning process of the market investors about the current prevailing volatility regime, their trading decision can be specified as a function of the continuously updated prior distribution of the states, which is shown in (VI.5). The addition of $\theta \Pr[S_{i2t} = 1 | \Psi_{t-1}]$ ensures the learning effects are taken into account. Particularly, when $S_{i2t} = 0$ and $\Pr[S_{i2t} = 1 | \Psi_{t-1}]$ is non-zero, the learning about an unanticipated high volatility regime occurs. On the other hand, if $S_{i2t} = 1$, the high volatility regime was fully anticipated *a priori*, $\mu_2 + \theta \Pr[S_{i2t} = 1 | \Psi_{t-1}]$ measures the increase in risk premium that the investors demand when facing an anticipated increases in volatility.

VI.4 Data

The data used in this study include monthly aggregate stock market returns for Australia, Canada, Sweden, Switzerland, the United Kingdom and the United States, obtained from *Global Financial Data*. The use of monthly data in our analysis would allow us to focus on the modelling of unconditional variance regime shifts rather than volatility clustering in a high-frequency sampled data, like the weekly data or daily data (see the discussion on motivation of utilising Markov switching model rather than GACRH models in section VI.3.2). Having said this, however, it would be interesting to see how the conditional variance behaves in each regime that Markov switching model has identified in high-frequency data samples. We leave this to future research. Table VI-1 presents the data sources and time span of the stock return data for each country. Table VI-2 presents the descriptive statistics for each country's return series. The returns in each market are calculated as the first difference of the natural logarithm of the stock market indices. As we can see from the table, the sample mean of all return series are greater than zero, with the highest in the Sweden market. Judging from the sample standard deviations, the Sweden market appears to be the most volatile one also. The null of normality of the return series has also been rejected significantly, as evident in the negative skewness of all the return series (except the UK) and positive excess kurtosis.

Table VI-1: Data sources and time span of the stock return data

Australia (AUS)	Canada (CAN)	Sweden (SWE)	Switzerland (SWI)	United Kingdom (UK)	United States (US)
ASX All ordinaries index	TSX composite	Stockholm Stock Exchange	SIX Swiss market index	LSE FTSE all shares index	New York Stock Exchange
January 1875 to September 2007	November 1917 to September 2007	December 1905 to September 2007	January 1916 to September 2007	January 1800 to September 2007	January 1800 to September 2007

Table VI-2: Descriptive statistics of the stock market return series for each country

	AUS	CAN	SWE	SWI	UK	US
Mean	0.0045	0.0045	0.0049	0.0032	0.0021	0.0026
Median	0.0054	0.0069	0.0061	0.0032	0.0020	0.0023
Maximum	0.2170	0.2059	0.2430	0.2878	0.5353	0.3524
Minimum	-0.4891	-0.3346	-0.3875	-0.2813	-0.3271	-0.3563
Std. Dev.	0.0382	0.0459	0.0480	0.0435	0.0375	0.0435
Skewness	-1.4677	-1.0352	-0.6388	-0.5678	1.6697	-0.5904
Kurtosis	24.2277	8.7602	8.6463	8.6395	36.8600	13.7562
Normality(JB)	30462.2 [0.000]	1687.5 [0.000]	1704.9 [0.000]	1527.8 [0.000]	120202.9 [0.000]	12157.9 [0.000]
Observations	1592	1081	1221	1108	2492	2492

A visual inspection of the stock market indices and return series for all countries are plotted in Figure VI-1 to Figure VI-8. All volatile movements of the return series are closely linked to global and local historical events. For example, the most easily identified two events that have caused extreme movements in all countries' stock markets are the 1933 great depression and the 1987 stock market crash. Other local historical events, linked to the UK stock market, are the banking crisis of 1825 and the second banking crisis of 1973-1975; and for the US stock market, are the American civil war and the Second World War. A snapshot of the more recent sample period for the UK and the US stock markets are plotted in Figure VI-6 and Figure VI-8, with the shaded areas represent the periods of recessions dated by NBER. The most volatile periods are tend to coincide with dated recession periods for both countries, which suggests that the recurrent appearance of high volatility regimes co-move with business cycles, despite the unidentified causal relations.

Figure VI-1: Australian stock market index and stock return

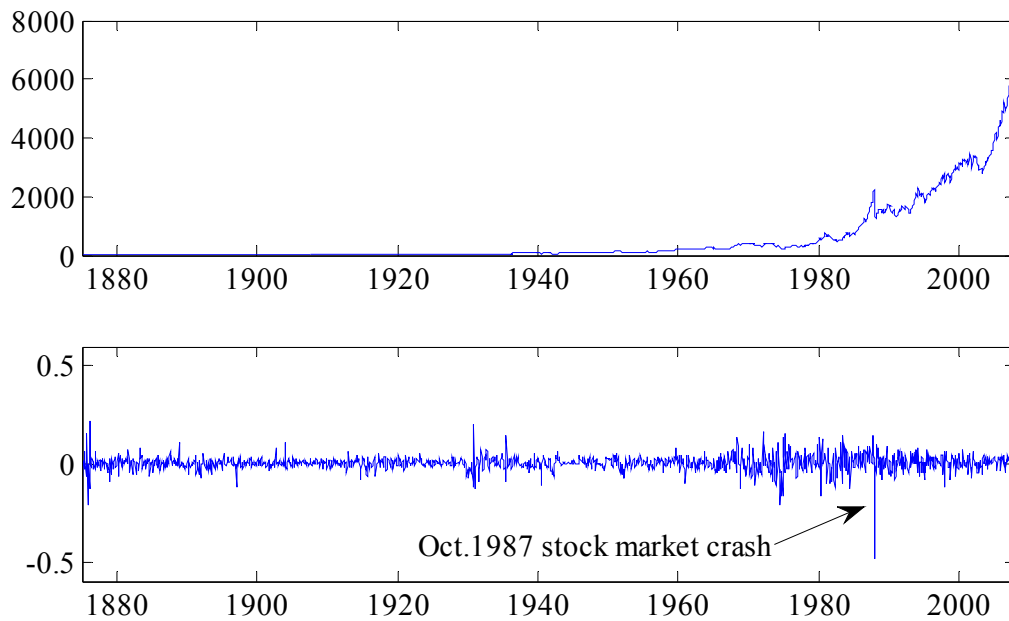


Figure VI-2: Canadian stock market index and stock return

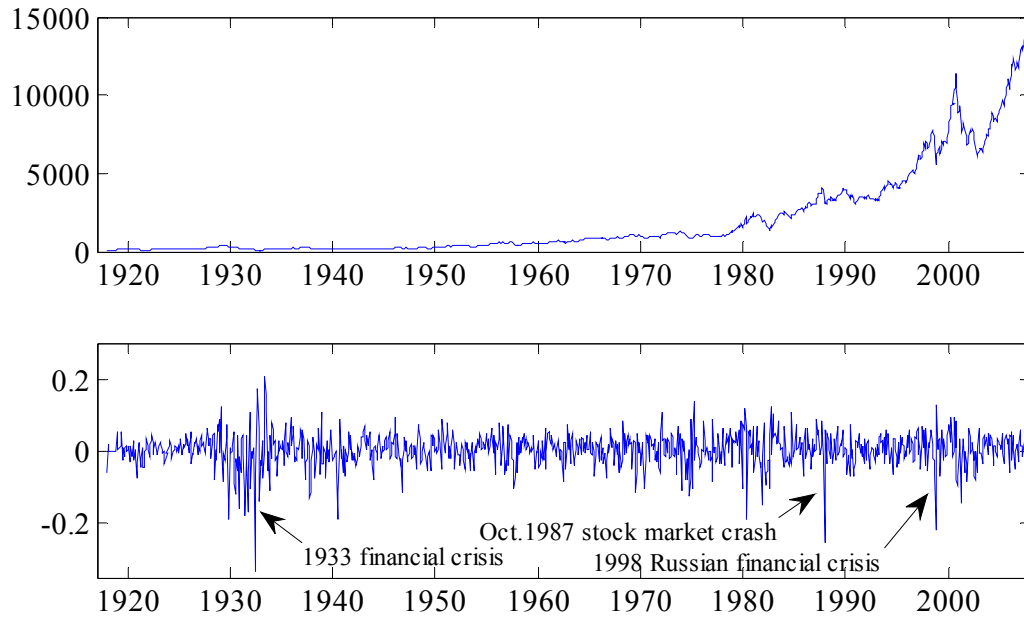


Figure VI-3: Swedish stock market index and stock return

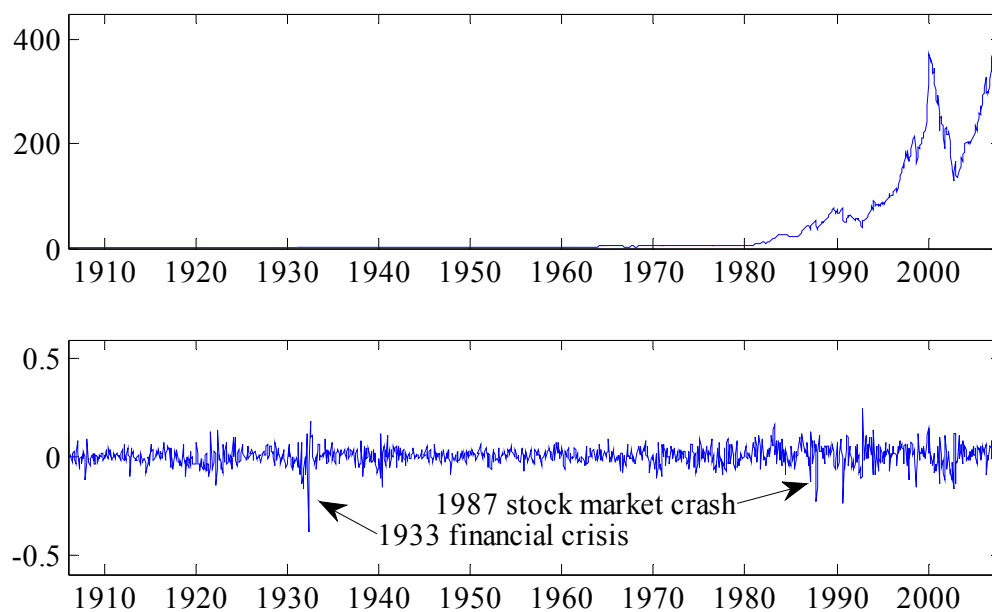


Figure VI-4: Switzerland stock market index and stock return

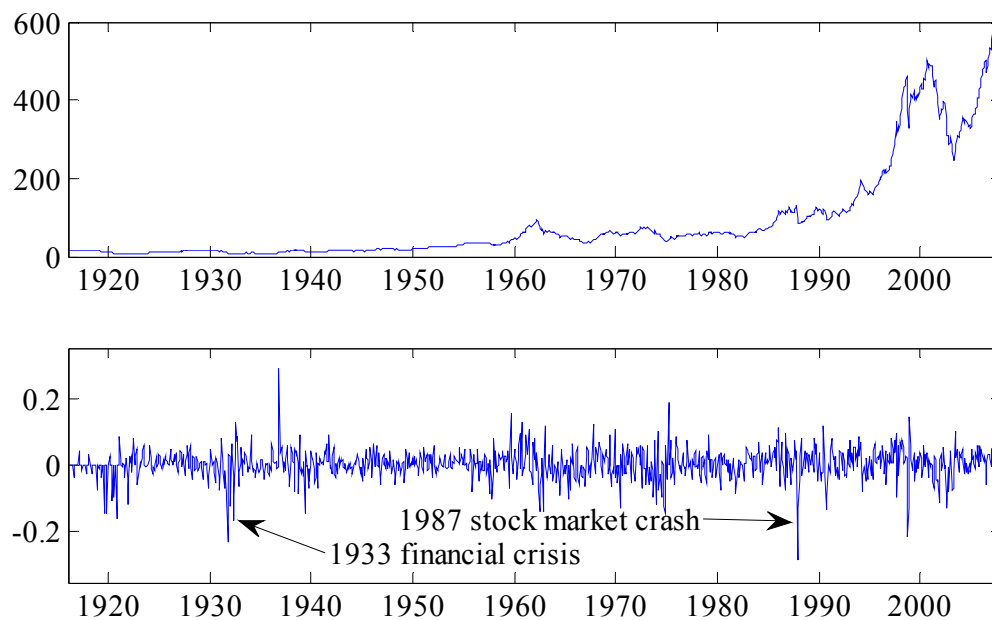


Figure VI-5: The UK stock market index and stock return

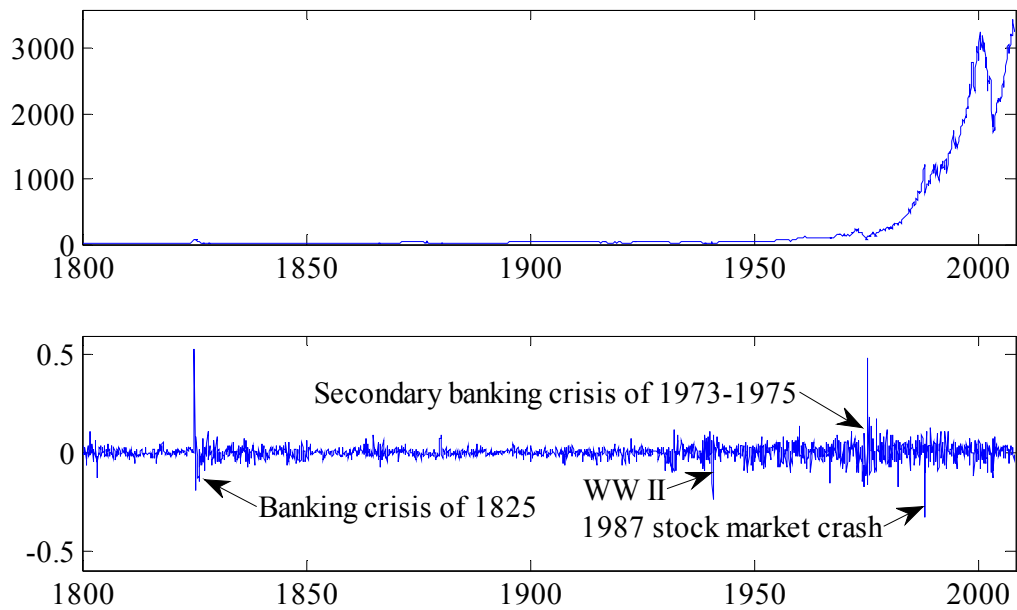


Figure VI-6: The UK stock market index and stock return (Sep. 1902 to Sep. 2007)

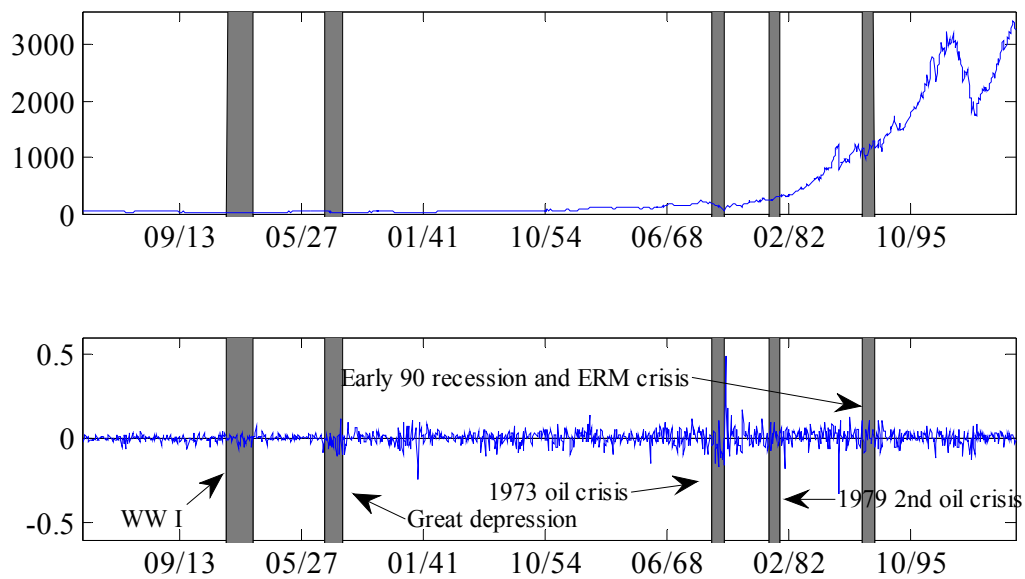


Figure VI-7: The US stock market index and stock return

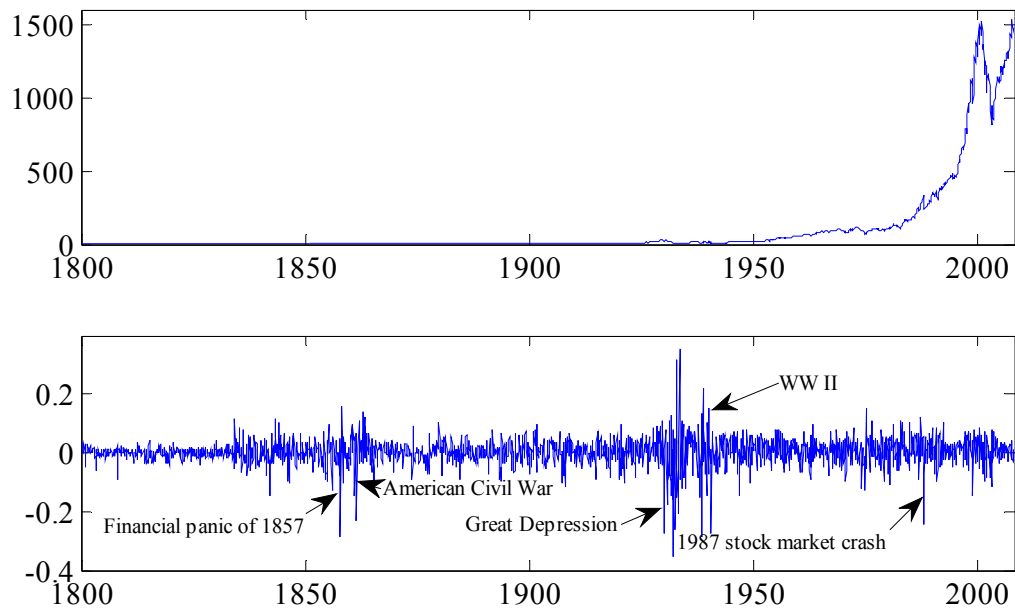
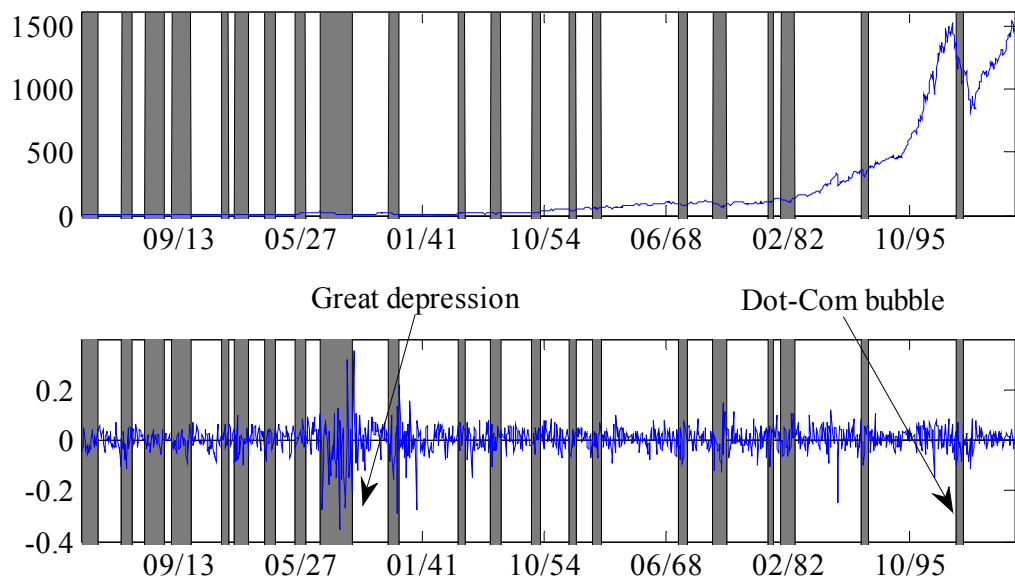


Figure VI-8: The US stock market index and stock return (Sep. 1902 to Sep. 2007)



VI.5 Estimation Results

In estimating the aforementioned models, we consider wide range of model specifications. First of all, we need to determine k_i - the number of regimes each stock return

series has. To do so, we undertake an extensive specification search and a series of tests, considering values of $k_i = 1, 2, 3$ and the order of the autoregressive term $p = 0, 1$ for each country in our data sample. First of all, it is important to know if the multiple regimes are necessary for the stock return data series considered. In other words, we begin from testing the linearity (single regime) of the return series against the alternative of having two distinct regimes in stock return for each country. It is well known that for Markov switching models the standard likelihood ratio test of the null hypothesis of linearity does not have the usual χ^2 distribution. The reason is that there are nuisance parameters which cannot be identified under the null hypothesis. As a result, the scores evaluated at the null hypothesis are identically zero. Hansen (1992) and Garcia (1998) introduce alternative tests of the linearity against multiple regimes. In this study, we use Hansen (1992) procedure, which provides an upper bound of the p -value for linearity, to determine the significance of improvement of allowing Markov-switching disturbance terms in the two components. In addition, we also consider more conventional ways of selecting models based on the Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (BIC). Finally, we verify our model selection results by running a series of residual diagnostic tests to see if the selected model could capture serial correlation and heteroskedasticity in the data series.

In order to implement Hansen (1992)'s procedure, we need to evaluate the constrained likelihood under the null hypothesis over a grid of values for the nuisance parameters. As noted in Hansen (1992), the only practical way to evaluate the maximal statistics is to form a grid search over a relatively small number of values of the nuisance parameters. A trade-off arises since a more extensive grid search means a major computational burden, but reduces the arbitrariness associated with the choice of grid. Define Model 1 as the restricted model under the null hypothesis of a single regime of the stock return's mean and volatility, the nuisance parameters in each alternative model specifications (Model 2 to 4) with $k_i = 2$ are $\{p_{i11}, p_{i22}, \sigma_{i2}\}$ in Model 2, $\{p_{i11}, p_{i22}, \sigma_{i2}, \mu_{i2}\}$ in Model 3 and $\{p_{i11}, p_{i22}, \sigma_{i2}, \mu_{i2}, \alpha_{i2}\}$ in Model 4, where i denotes the country index. To select a proper grid for these nuisance parameters, we follow Kim, Morley and Piger (2005) to consider a grid covering the likely values of these when estimating the Markov switching alternative. The grids that we used for p_{i11} and p_{i22} are $[0.35, 0.99]$, with increment step of 0.01. The grids for σ_{i2} and μ_{i2} are

$[0.005, 0.2]$ and $[0.005, 0.1]$, respectively, each in increment step of 0.005. The grids for α_{i2} in Model 5 range from 0.1 to 0.4 in increment step of 0.05.

The results of the Hansen test, which are implemented as described above, are shown in Table VI-3. The conservative upper bound p-values for linearity obtained are all zeros for every model considered and for all countries in our sample. The results clearly indicate a strong rejection of linearity and favour the multi-regime setting for stock returns' first and/or second moments.

Table VI-3: Model specification test using Hansen 1992 test procedure

	Model 2	Model 3	Model 4
AUS	98.448 [0.000]	101.420 [0.000]	107.871 [0.000]
CAN	157.364 [0.000]	158.412 [0.000]	129.356 [0.000]
SWE	94.067 [0.000]	124.425 [0.000]	77.137 [0.000]
SWI	236.301 [0.000]	195.847 [0.000]	108.387 [0.000]
UK	119.531 [0.000]	136.221 [0.000]	120.630 [0.000]
US	216.644 [0.000]	220.309 [0.000]	186.455 [0.000]

Note: The null hypothesis of this linearity test is defined as a single regime model as in Model 1. In alternative hypothesis, stock return's mean and volatility are allowed to switch between 2 regimes - the low and high volatility regimes. The Hansen 1992 test's conservative p-values are reported in square brackets.

Having rejected the single regime model of the stock returns for each country, we now turn to the estimation of multi-regime models. For each model (Model 2, 3 and 4), we test the significance of the inclusion of the autoregressive term. In addition, we assume the underlying states in each model are either 2 or 3. Since Model 2 and 3 are nested in Model 4, and the case when having $k_i = 2$ are nested under $k_i = 3$, the "best" in-sample performing model for each country is determined by the information criteria (AIC, BIC and HQIC). To verify the model selection results, we also perform a residual diagnostic test for each model. Specifically, we test the overall randomness of the residuals of the models under the null hypothesis of randomness. We report two Ljung-Box Q statistics for each model: one is the autocorrelation Q statistics based on the standardized residuals up to 12 lags; the other one is the ARCH effect Q^2 statistics based on the squared standardized residuals up to 12 lags.

VI.5.1 Australia

Table VI-4 reports the parameter estimates of models 1, 2, 3 and 4 on Australian stock market returns. Along with the parameter estimates, we also report the values of information criteria that will be used in determining the preferred model, based on the trade-off between model's goodness-of-fit and parsimony. The upper panels (A1 and A2) assume there are two ($k = 2$) underlying regimes in the stock return's expected mean or/and unconditional variances, while the lower panels (B1 and B2) assume there are three regimes ($k = 3$). The left panels (A1 and B1) consider the case where the autoregressive term is not included (AR(0)), while the right panels (A2 and B2) take this term into account (AR(1)) to proxy the missing variables that may track the time-varying expected mean of the stock returns.

Looking at panels A1 and A2, models with AR(1) are always preferred to models with AR(0), according to what the three information criteria suggest; and Model 1 is strongly rejected in favour of 2-regime models. Adding an extra parameter, μ_2 , to allow a regime dependent expected return in Australian stock returns does not deepens BIC and HQ, but is preferred according to AIC. However, the insignificant estimates of μ_2 in Model 3-AR(1) and Model 4-AR(1) doubt the regime dependent specification on the expected return. The Ljung-Box statistics shown in Table VI-5 suggest that, under a 2-regime specification, Model 2-AR(1) is better than Model 3-AR(1) in capturing the heteroscedasticity effect in the residuals, while the serial correlation problem disappears once the autoregressive term is included in all models.

Table VI-4: Estimation results (Australia)

$k_{AUS} = 2$	Panel A1: AR(0)			Panel A2: AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$	0.00448 (0.00096)	0.00568 (0.00064)	0.00604 (0.00067)	0.00453 (0.00102)	0.00574 (0.00069)	0.00610 (0.00072)	0.00613 (0.00074)
μ_2			0.00116 (0.00280)			0.00101 (0.00307)	0.00093 (0.00293)
σ_1	0.03819 (0.00068)	0.01985 (0.00067)	0.01987 (0.00065)	0.03810 (0.00068)	0.01968 (0.00069)	0.01969 (0.00068)	0.01966 (0.00070)
σ_2		0.06090 (0.00238)	0.06090 (0.00237)		0.06102 (0.00243)	0.06102 (0.00243)	0.06091 (0.00244)
p_{11}		0.96731 (0.00797)	0.96702 (0.00799)		0.96553 (0.00851)	0.96511 (0.00855)	0.96452 (0.00878)
p_{22}		0.93168 (0.01741)	0.93037 (0.01773)		0.92683 (0.01897)	0.92529 (0.01935)	0.92423 (0.01979)
α_1				0.05927 (0.02500)	0.08913 (0.02528)	0.08856 (0.02525)	0.11361 (0.03215)
α_2							0.03742 (0.04666)
Log L	2939.166	3273.214	3274.617	2941.205	3278.636	3279.911	3280.762
AIC	-3.68991	-4.10580	-4.10630	-3.69353	-4.11394	-4.11428	-4.11409

BIC	-3.68315	-4.08892	-4.08605	-3.68340	-4.09368	-4.09065	-4.08708
HQ	-3.67987	-4.09953	-4.09878	-3.68978	-4.10641	-4.10550	-4.10406
$k_{AUS} = 3$	Panel B1: AR(0)			Panel B2: AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$		0.00572 (0.00063)	0.00576 (0.00067)		0.00555 (0.00068)	0.00596 (0.00077)	0.00604 (0.00079)
μ_2			0.00441 (0.00206)			0.00366 (0.00184)	0.00373 (0.00184)
μ_3			-0.01824 (0.02343)			-0.00367 (0.01588)	-0.00525 (0.01291)
σ_1		0.01848 (0.00074)	0.01859 (0.00069)		0.01700 (0.00069)	0.01693 (0.00068)	0.01690 (0.00067)
σ_2		0.04439 (0.00509)	0.04435 (0.00269)		0.03953 (0.00185)	0.03942 (0.00186)	0.03936 (0.00181)
σ_3		0.06138 (0.00778)	0.12488 (0.01692)		0.11653 (0.01304)	0.11571 (0.01301)	0.11326 (0.01215)
p_{11}		0.88486 (0.03813)	0.97160 (0.00748)		0.96184 (0.01055)	0.96053 (0.01051)	0.96063 (0.01052)
p_{12}		0.11514 (0.03813)	0.01321 (0.00681)		0.03733 (0.01102)	0.03947 (0.01051)	0.03937 (0.01052)
p_{21}		0.86466 (0.05958)	0.01598 (0.01014)		0.04237 (0.01316)	0.04532 (0.01284)	0.04433 (0.01296)
p_{22}		1.3E-133 (1.7E-127)	0.96418 (0.01225)		0.93869 (0.01538)	0.93495 (0.01512)	0.93576 (0.01528)
p_{31}		0.03256 (0.02659)	0.41613 (0.20627)		0.01868 (0.03163)	2.1E-30 (1.4E-24)	0.01356 (0.02598)
p_{32}		0.00071 (0.02987)	0.17608 (0.09628)		0.17286 (0.06581)	0.18979 (0.06588)	0.18917 (0.06934)
α_1					0.12879 (0.02503)	0.12788 (0.02514)	0.15096 (0.03948)
α_2							0.13022 (0.03638)
α_3							-0.14206 (0.15393)
Log L		3287.643	3321.854		3341.060	3341.764	3343.833
AIC		-4.11764	-4.15811		-4.18612	-4.18449	-4.18458
BIC		-4.08389	-4.11761		-4.14898	-4.14060	-4.13393
HQ		-4.10511	-4.14307		-4.17233	-4.16819	-4.16577

Note: the standard errors are in parentheses. The standard errors are numerically calculated using the variance-covariance matrix of the estimators (see e.g. Hamilton (1994, p.389) for details).

Table VI-5: Ljung-Box Q statistics for autocorrelation effect and Q^2 statistics for ARCH effect on the residuals (Australia)

	2-regime AR(0)		2-regime AR(1)		3-regime AR(0)		3-regime AR(1)	
	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$
Model 2	28.837 [0.004]	15.402 [0.220]	19.832 [0.070]	21.300 [0.046]	32.990 [0.001]	9.111 [0.693]	19.662 [0.074]	14.765 [0.255]
Model 3	28.242 [0.005]	17.203 [0.142]	19.796 [0.071]	23.552 [0.023]	24.369 [0.018]	16.389 [0.174]	20.150 [0.064]	15.163 [0.233]
Model 4			19.542 [0.076]	25.499 [0.013]			18.949 [0.090]	16.571 [0.166]

Note: This Ljung-Box Q statistics is a modified portmanteau test of the original one, which allows us to test the ARCH effect of the residuals. Q(12) refers to the Ljung-Box test for no serial correlation up to 12 lags. $Q^2(12)$ refers to the Ljung-Box test for no ARCH effect up to 12 lags. The p-values are shown in square brackets. A smaller p-value indicates a rejection of the null hypothesis of no serial correlation or no ARCH effect.

Each of the two regimes identified in Model 2-AR(1) has a clear economic interpretation. The first regime, which has a low monthly volatility of 1.968%, can be interpreted as a calm market, while the second regime, which has a much higher monthly volatility of 6.102% can be interpreted as a volatile market. The expected duration of the calm market is 29 months⁴¹, which is more persistent than the volatile market that has an expected duration of 14 months. To further assist with the economic interpretation of the two regimes identified, Figure VI-9 plots the smoothed state probability from Model 2-AR(1). The top figure is a time series plot of the Australian monthly stock returns from March 1875 to September 2007. The bottom figure is drawn assuming each month's return is taken from one of the two normal densities with different variances but same mean according to an unobserved Markov chain process. Since the market investors in Model 2-AR(1) are assumed to observe this Markov chain perfectly but not the econometricians, the figure plots the econometrician's probability that the stock return in that month is drawn from a high variance density. There is a clear matching between the probability of observing a high volatility regime and many well-known historical events, e.g. the 1929-1933 Great Depression, the 1973 first oil crisis, the 1979 second oil crisis and the 1987 stock market crash. We don't report the smoothed state probability for other models, but the similarity of the state probability across models can be seen from Panel A of Table VI-6, where we show the correlation between the smoothed state probability of observing a high volatility regime across models.

⁴¹ calculated as $1/(1 - p_{11})$

Figure VI-9: Australian stock market return series and Model 2-AR(1)'s smoothed posterior probability of observing a high volatility regime (2-regime)

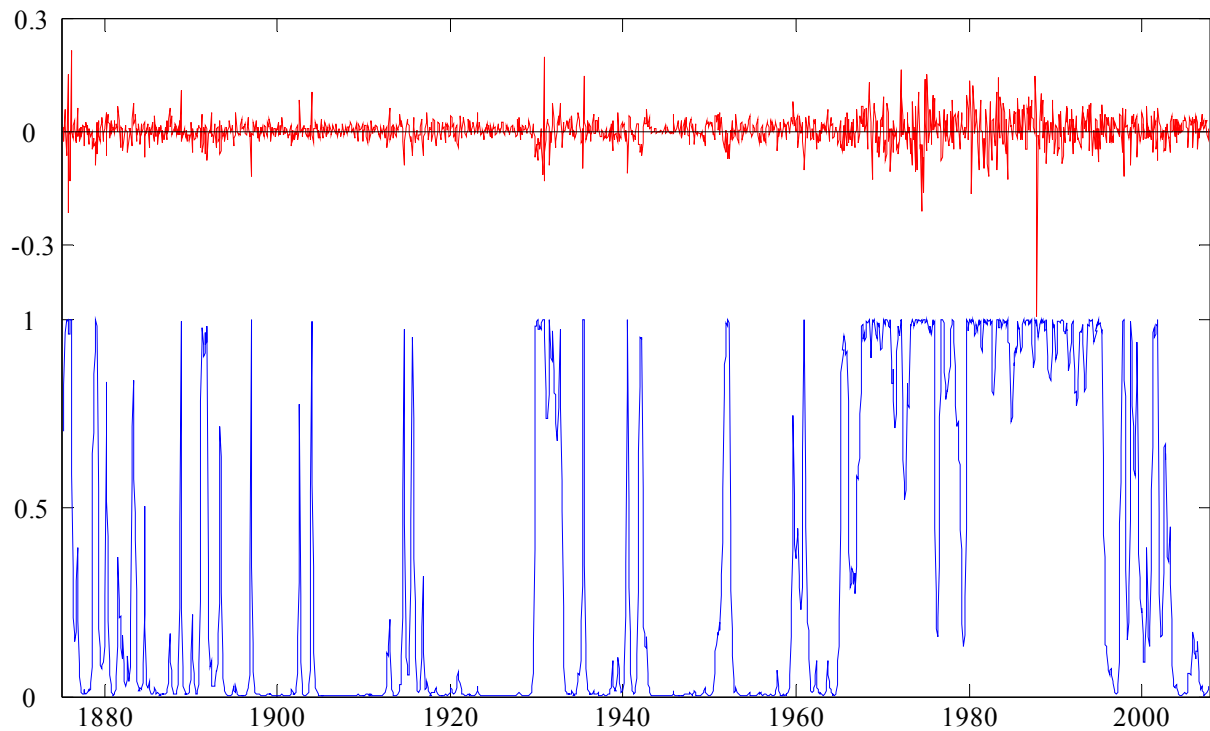


Table VI-6: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - Australia

	Model 2 AR(0)	Model 2 AR(1)	Model 3 AR(0)	Model 3 AR(1)	Model 4 AR(1)
Panel A: 2-regime (high volatility regime)					
Model 2 AR(0)	1	0.9974535	0.9995886	0.9965061	0.9940285
Model 2 AR(1)		1	0.9977058	0.9996483	0.9989969
Model 3 AR(0)			1	0.9974955	0.99526
Model 3 AR(1)				1	0.9995882
Model 4 AR(1)					1
Panel B: 3-regime (low volatility regime)					
Model 2 AR(0)	1	0.8298813	0.895198	0.8301513	0.8264662
Model 2 AR(1)		1	0.9408622	0.9984188	0.9990431
Model 3 AR(0)			1	0.9350281	0.9336026
Model 3 AR(1)				1	0.9990103
Model 4 AR(1)					1
Panel C: 3-regime (high volatility regime)					
Model 2 AR(0)	1	0.4330569	0.2307493	0.4416208	0.4352588
Model 2 AR(1)		1	0.7077076	0.999361	0.9993027
Model 3 AR(0)			1	0.6942537	0.701896
Model 3 AR(1)				1	0.9988566
Model 4 AR(1)					1

Now, consider the results of models assuming a 3-regime specification in panel B1 and B2 of Table VI-4. Generally, all 3-regime models are preferred to their counterparts with a 2-regime specification. Given an additional regime, models are more capable to capture a negative expected mean in stock returns. While all three information criteria prefer Model 2-AR(1) as the best in-sample performing model under a 3-regime specification, the Ljung-Box statistics in Table VI-5 suggest that Model 4-AR(1) is more able to capture serial correlation in the residuals than Model 2-AR(1). However, the high standard errors of the estimates μ_3 and α_3 suggest that Model 4-AR(1) may face the risk of over-parameterization.

Figure VI-10: Australian stock market return series and Model 2-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)

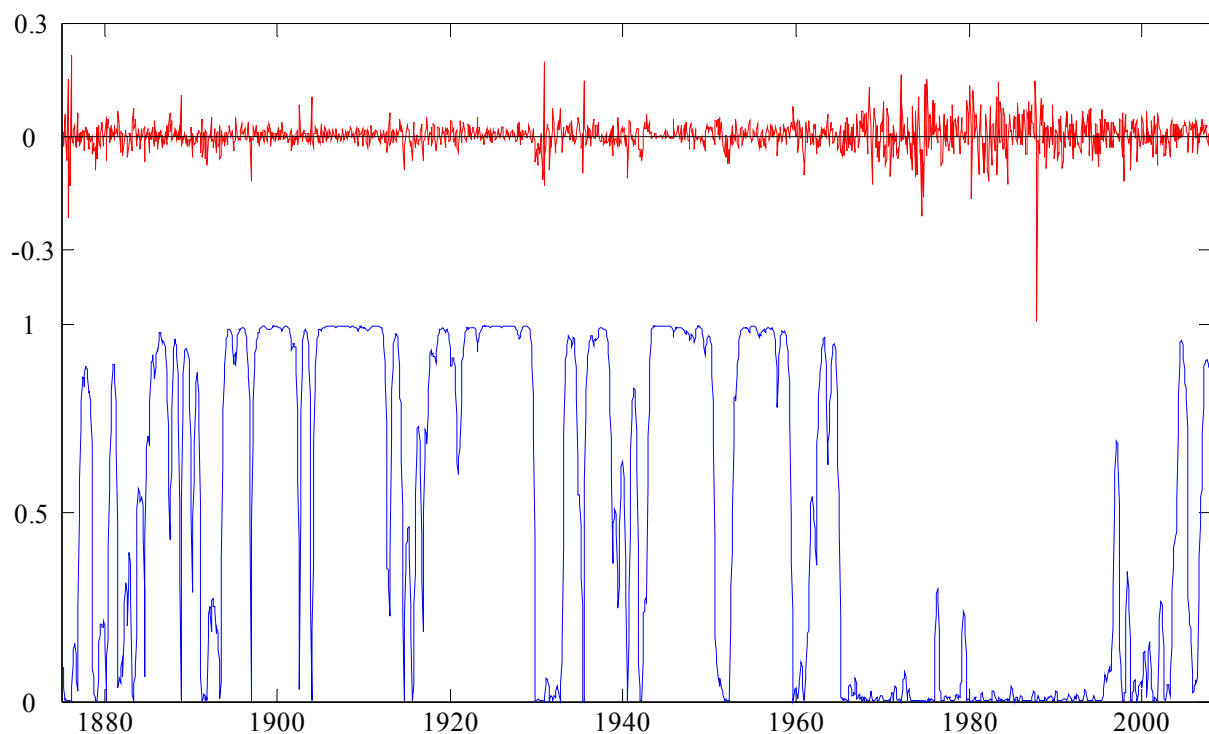
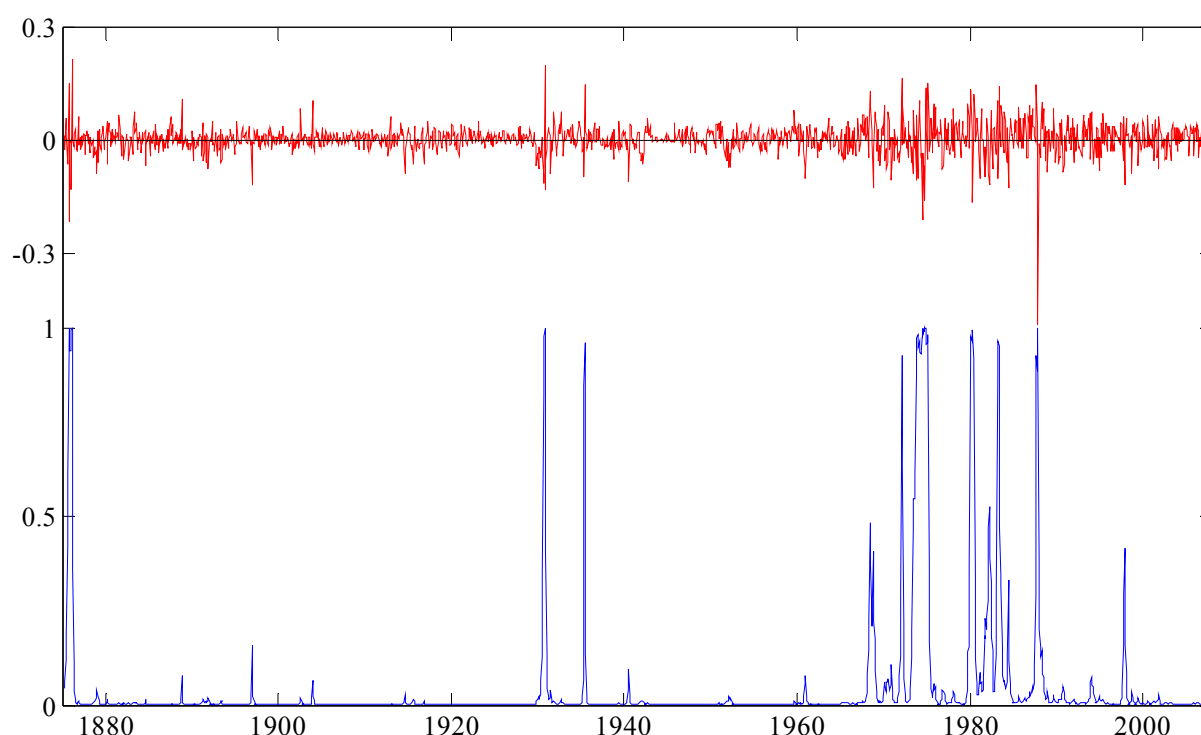


Figure VI-10 and Figure VI-11 show the smoothed state probability of observing a low and high volatility regime, respectively, from Model 2-AR(1) under a 3-regime specification. A comparison between Figure VI-11 and Figure VI-9 reveals that the high volatility regime under the 3-regime framework captures the extreme movements of the stock market which are short-lived. This helps us to capture the stock return's leptokurtosis in its distribution. In addition, as most of the extreme movements in the stock market are downward drops, Markov switching models with 3 regimes would better capture the negative skewness in the stock returns than their 2-regime counterparts. The additional regime

considered in a 3-regime specification can be thought as the recovery regime from a high volatile state to a calm state. The monthly unconditional volatility in this regime is around 4%, which is far less than the extreme volatile state with a 11.6% monthly volatility but higher than the calm state that has a monthly volatility of 1.7%.

Figure VI-11: Australian stock market return series and Model 2-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)



The correlations between the smoothed state probabilities of observing low and high volatility regimes, across different models under the 3-regime specification, are shown in Panel B and C in Table VI-6. A significant difference between the correlation coefficients in Panel C and those in Panel A is that the coherence of the probability of observing a high volatility regime under a 3-regime framework has decreased in general, comparing to that under a 2-regime framework. More specifically, the autoregressive term under the 3-regime framework plays a more important role than under a 2-regime specification. A clear pattern shows that models with the autoregressive term share greater similarity in capturing the high volatility regime while the models without this autoregressive term are seemed disconnected with other models. For example, the correlation between Model 2-AR(0) and Model 3-AR(0) is merely 0.23. Having in mind that the difference between these two models is the extra dynamic on the expected mean that we added into Model 3-AR(0), the low correlation between the probability of observing a high volatility regime from these two models must arise from the different parameterizations of the expected return. Model 3-AR(0) specifies the

expected return of the stock market to be a regime dependent time-varying process. Although the negative expected return reported in Panel B1 of Table VI-4 is not statistically significant, this 3-regime specification of the expected return notably linked the negative return to a higher volatility regime implying a time-varying systematic risk in the stock market which are predictable. Model 2-AR(0), however, assumes that an equally weighted portfolio (like the aggregate stock market index in our case) should have a constant expected return. In other words, the recurrent shifts in the volatility regime should not affect the systematic risk component if the portfolio is well diversified. In absence of the regime dependent expected return, the market volatility captures not only the predictable shifts in systematic risks but also the idiosyncratic risk. Therefore, we would expected a longer duration of high volatility regime in Model 2-AR(0) than in Model 3-AR(0). In fact, the expected duration of high volatility regime calculated for Model 2-AR(0) and Model 3-AR(0) are 30 and 1.6 months, respectively. Visual inspection from Figure VI-12 and Figure VI-13 further confirm this.

Figure VI-12: Australian stock market return series and Model 2-AR(0)'s smoothed posterior probability of observing a high volatility regime (3-regime)

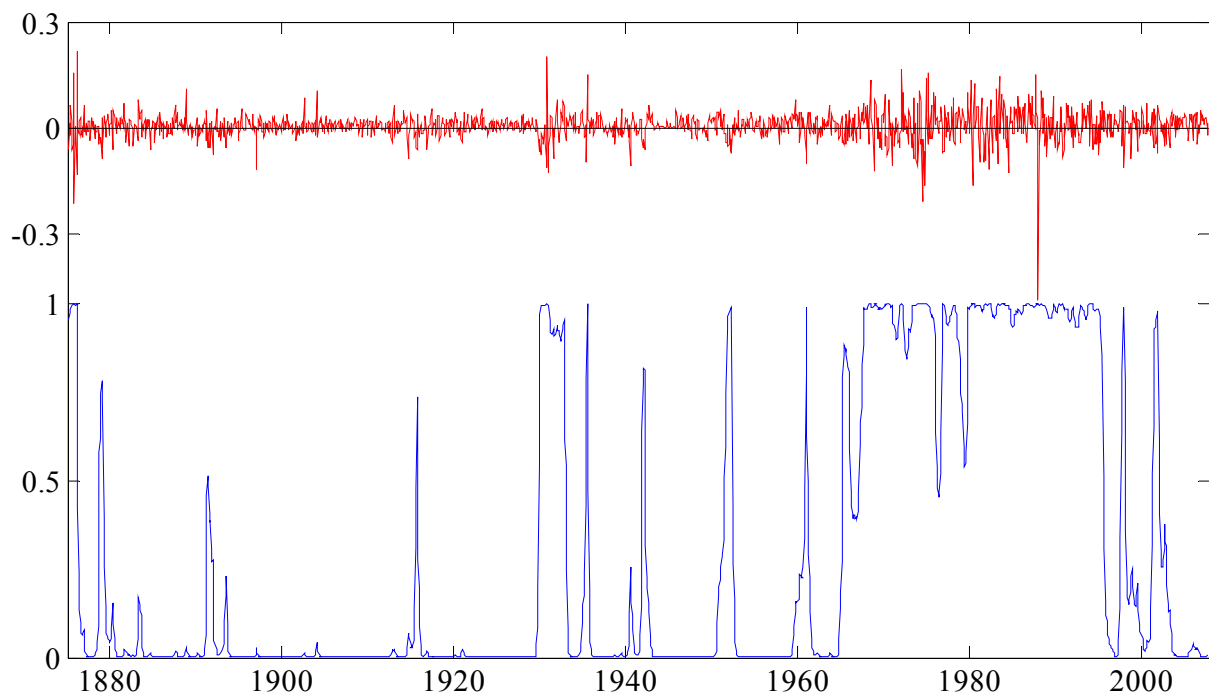
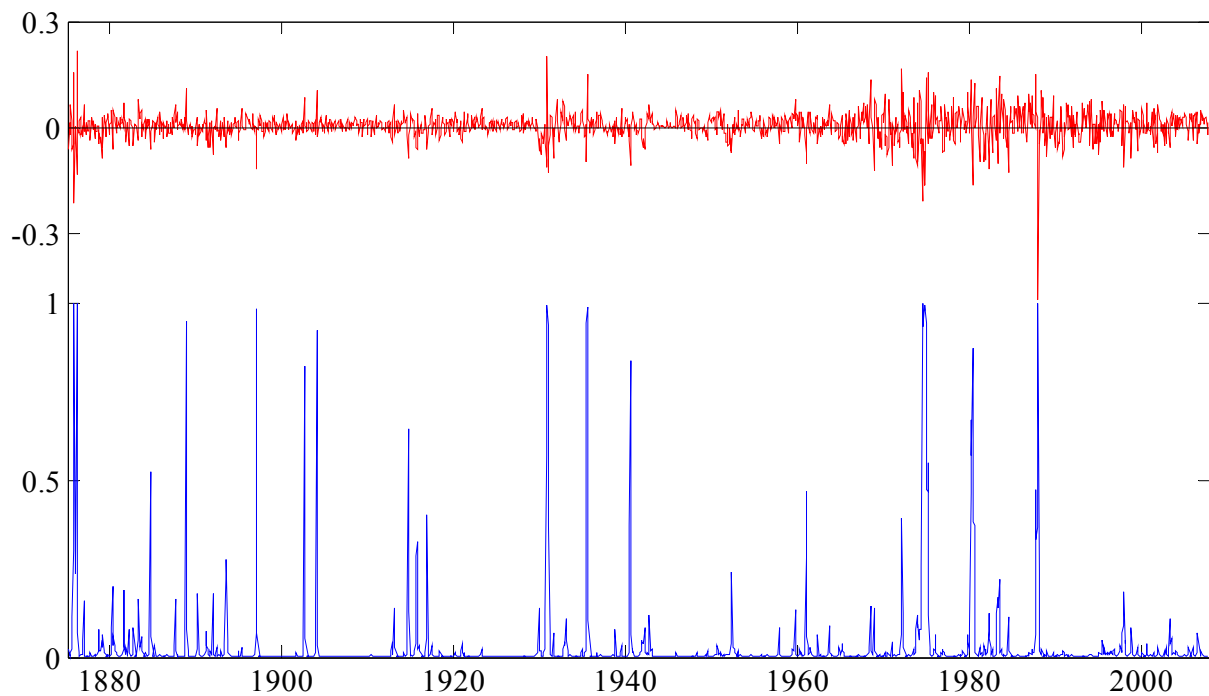


Figure VI-13: Australian stock market return series and Model 3-AR(0)'s smoothed posterior probability of observing a high volatility regime (3-regime)



VI.5.2 Canada

Table VI-7 reports the parameter estimates for various models on Canadian stock market returns. Unlike the Australian case, Canadian stock market strongly demands a regime dependent expected return. For the preferred Model 3-AR(1) under a 2-regime specification, the second regime identified in the Canadian market is a bear market state with a negative expected mean of -1.523% but has a very high monthly volatility of 7.946%. Figure VI-14 plots the smoothed state probability of observing a high volatility regime from Model 3-AR(1). There is a clear matching between the probability of observing a high volatility regime and many well-known historical events, e.g. the 1929-1933 Great Depression, the 1973 first oil crisis, the 1979 second oil crisis, the 1987 stock market crash and the 1998 Russian financial crisis. Notably, these historical events are always associated with large negative drops of the Canadian stock market returns. We don't report the smoothed state probability for other models, but the similarity of the state probability across models can be seen from Panel A of Table VI-9, where we show the correlation between the smoothed state probabilities of observing a high volatility regime across models.

Table VI-7: Estimation result (Canada)

$k_{AUS} = 2$	AR(0)			AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$	0.00442 (0.00140)	0.00748 (0.00114)	0.00886 (0.00120)	0.00449 (0.00167)	0.00739 (0.00129)	0.00874 (0.00133)	0.00878 (0.00130)
μ_2			-0.01673 (0.00692)			-0.01523 (0.00729)	-0.01586 (0.00802)
σ_1	0.04587 (0.00099)	0.03248 (0.00104)	0.03211 (0.00106)	0.04513 (0.00097)	0.03221 (0.00109)	0.03184 (0.00109)	0.03182 (0.00108)
σ_2		0.08550 (0.00610)	0.08167 (0.00551)		0.08172 (0.00594)	0.07946 (0.00543)	0.07927 (0.00536)
p_{11}		0.97810 (0.00776)	0.97451 (0.00859)		0.97913 (0.00798)	0.97551 (0.00880)	0.97488 (0.00898)
p_{22}		0.89132 (0.03791)	0.87769 (0.03986)		0.90292 (0.03731)	0.88819 (0.03952)	0.88464 (0.04068)
α_1				0.17682 (0.02997)	0.12922 (0.03152)	0.11558 (0.03131)	0.09310 (0.03504)
α_2							0.20994 (0.07723)
Log L	1792.664	1924.149	1931.233	1808.557	1932.129	1937.664	1938.572
AIC	-3.32220	-3.56057	-3.57186	-3.35294	-3.57684	-3.58526	-3.58509
BIC	-3.31295	-3.53746	-3.54413	-3.33906	-3.54909	-3.55288	-3.54809
HQ	-3.30788	-3.55182	-3.56136	-3.34768	-3.56633	-3.57300	-3.57108
$k_{AUS} = 3$	AR(0)			AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$		0.00717 (0.00115)	0.00602 (0.00323)		0.00775 (0.00126)	0.01234 (0.00231)	0.01263 (0.00245)
μ_2			0.00967 (0.00140)			0.00709 (0.00187)	0.00714 (0.00182)
μ_3			-0.00346 (0.00709)			-0.01769 (0.00874)	-0.01860 (0.00963)
σ_1		0.02288 (0.00206)	0.02001 (0.00262)		0.01877 (0.00188)	0.01974 (0.00184)	0.01982 (0.00185)
σ_2		0.03435 (0.00320)	0.03532 (0.00392)		0.03461 (0.00277)	0.03566 (0.00262)	0.03555 (0.00263)
σ_3		0.08814 (0.00739)	0.08450 (0.00732)		0.08550 (0.00680)	0.08466 (0.00655)	0.08419 (0.00649)
p_{11}		0.99621 (0.00689)	0.92169 (0.05186)		0.97137 (0.02970)	0.94720 (0.02900)	0.94645 (0.02862)
p_{12}		0.00194 (0.00987)	0.06952 (0.04867)		0.02863 (0.02970)	0.05280 (0.02900)	0.05355 (0.02862)
p_{21}		0.00069 (0.00304)	0.01725 (0.01094)		0.00507 (0.00558)	0.01380 (0.01132)	0.01412 (0.01123)
p_{22}		0.97689 (0.00897)	0.96517 (0.01332)		0.97469 (0.00970)	0.96324 (0.01439)	0.96162 (0.01484)
p_{31}		0.00266 (0.01210)	0.00845 (0.01253)		3.5E-06 (0.00023)	1.7E-76 (8.3E-71)	2.1E-24 (2.7E-18)
p_{32}		0.10803 (0.04381)	0.08623 (0.03278)		0.09916 (0.03767)	0.11127 (0.04176)	0.11672 (0.04375)
α_1				0.13059 (0.03151)	0.12019 (0.03161)		0.16386 (0.10065)
α_2							0.08896 (0.03969)
α_3							0.22034 (0.08622)
Log L		1934.648	1936.289		1946.003	1952.328	1953.309
AIC		-3.57078	-3.57011		-3.59332	-3.60135	-3.59946
BIC		-3.52455	-3.51464		-3.54244	-3.54122	-3.53007
HQ		-3.55327	-3.54911		-3.57405	-3.57858	-3.57318

Note: the standard errors are in parentheses. The standard errors are numerically calculated using the variance-covariance matrix of the estimators (see e.g. Hamilton (1994, p.389) for details).

Table VI-8: Ljung-Box Q statistics for autocorrelation effect and Q^2 statistics for ARCH effect on the residuals (Canada)

	2-regime AR(0)		2-regime AR(1)		3-regime AR(0)		3-regime AR(1)	
	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$
Model 2	33.216 [0.001]	7.602 [0.815]	15.057 [0.238]	7.645 [0.812]	34.998 [0.001]	7.075 [0.853]	14.423 [0.275]	6.731 [0.875]
Model 3	29.566 [0.003]	9.904 [0.624]	15.917 [0.195]	9.416 [0.667]	30.662 [0.002]	5.943 [0.919]	13.459 [0.337]	7.212 [0.843]
Model 4			15.186 [0.231]	9.542 [0.656]			13.050 [0.365]	7.425 [0.828]

Note: This Ljung-Box Q statistics is a modified portmanteau test of the original one, which allows us to test the ARCH effect of the residuals. Q(12) refers to the Ljung-Box test for no serial correlation up to 12 lags. $Q^2(12)$ refers to the Ljung-Box test for no ARCH effect up to 12 lags. The p-values are shown in square brackets. A smaller p-value indicates a rejection of the null hypothesis of no serial correlation or no ARCH effect.

Figure VI-14: Canadian stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a high volatility regime (2-regime)

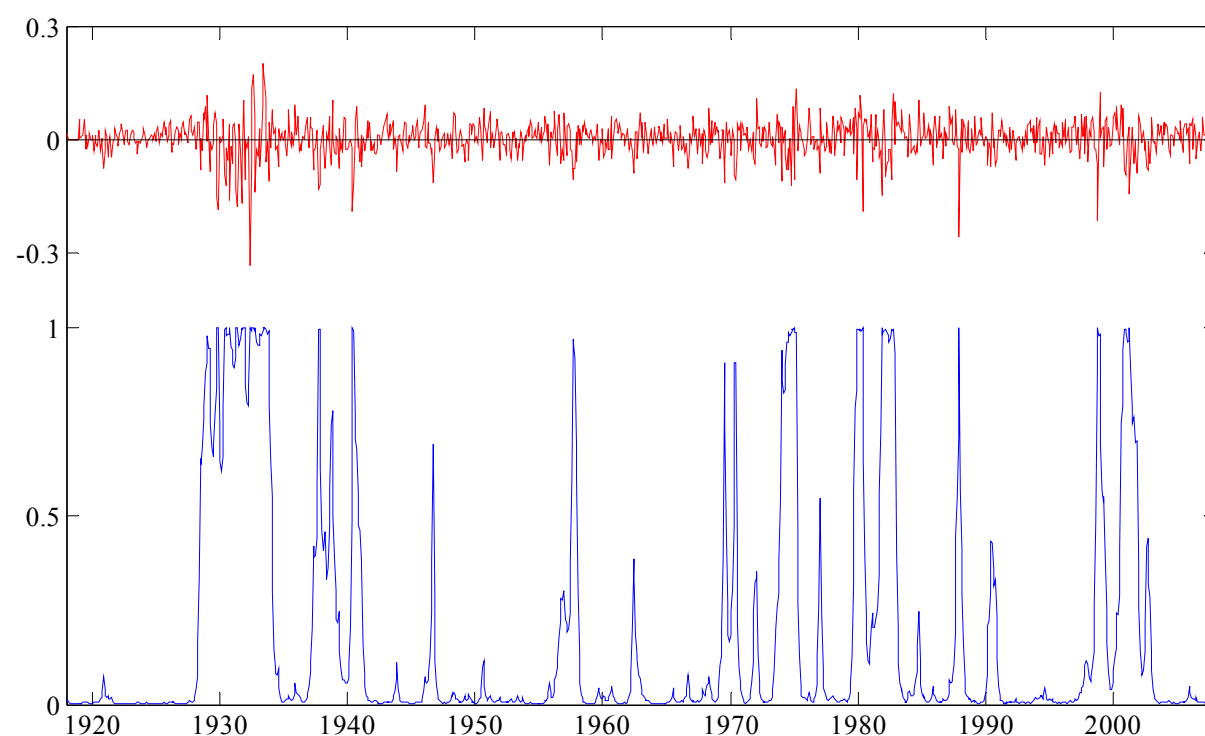


Table VI-9: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - Canada

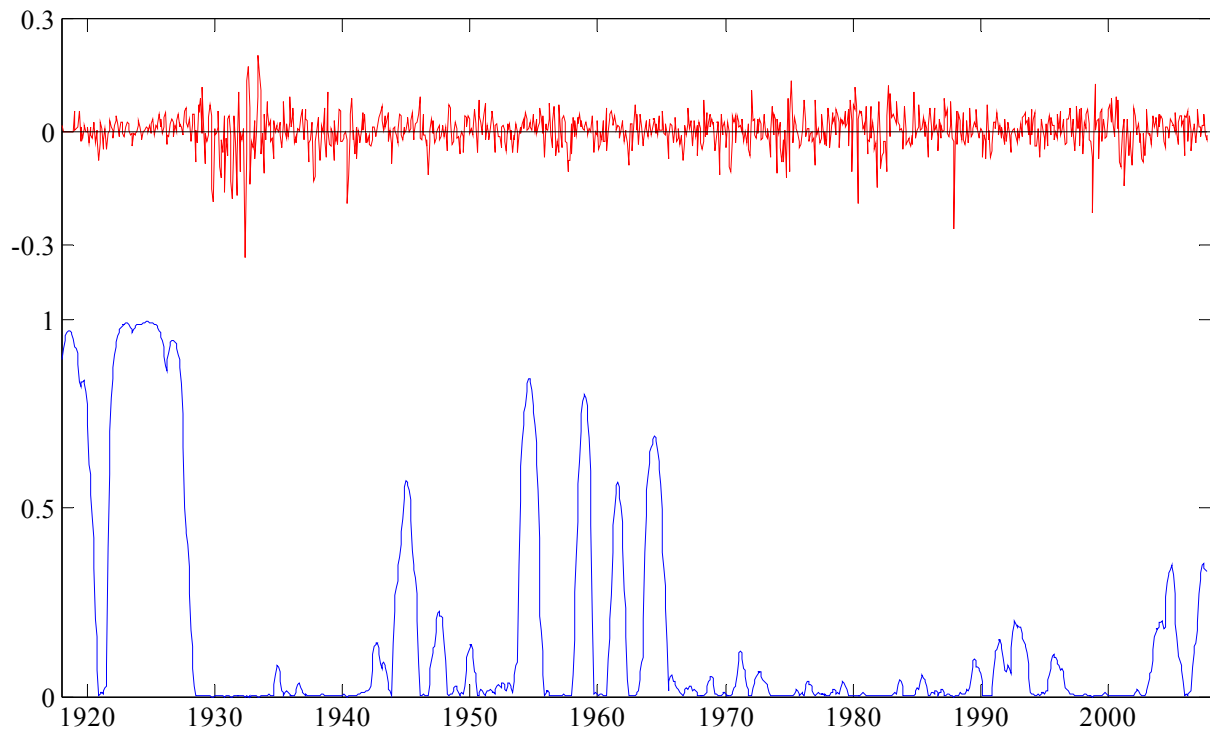
	Model 2 AR(0)	Model 2 AR(1)	Model 3 AR(0)	Model 3 AR(1)	Model 4 AR(1)
	2-regime (high volatility regime)				
Model 2 AR(0)	1	0.9931366	0.9906072	0.9904956	0.99218134

Model 2 AR(1)	1	0.9807142	0.9920125	0.99010395
Model 3 AR(0)		1	0.9945527	0.99706703
Model 3 AR(1)			1	0.99937199
Model 4 AR(1)				1
3-regime (low volatility regime)				
Model 2 AR(0)	1	0.9201393	0.810749	0.7684679
Model 2 AR(1)		1	0.929987	0.9037516
Model 3 AR(0)			1	0.9098898
Model 3 AR(1)				1
Model 4 AR(1)				1
3-regime (high volatility regime)				
Model 2 AR(0)	1	0.9939046	0.9949106	0.99349
Model 2 AR(1)		1	0.9907637	0.9955007
Model 3 AR(0)			1	0.9938369
Model 3 AR(1)				1
Model 4 AR(1)				1

In a 3-regime specification (where the parameter estimation results are shown in panel B1 and B2 of Table VI-7), all three information criteria select Model 3-AR(1) as the best in-sample performing model. The serial correlation and ARCH effect in the residuals are also captured by Model 3-AR(1), as shown in Note: the standard errors are in parentheses. The standard errors are numerically calculated using the variance-covariance matrix of the estimators (see e.g. Hamilton (1994, p.389) for details).

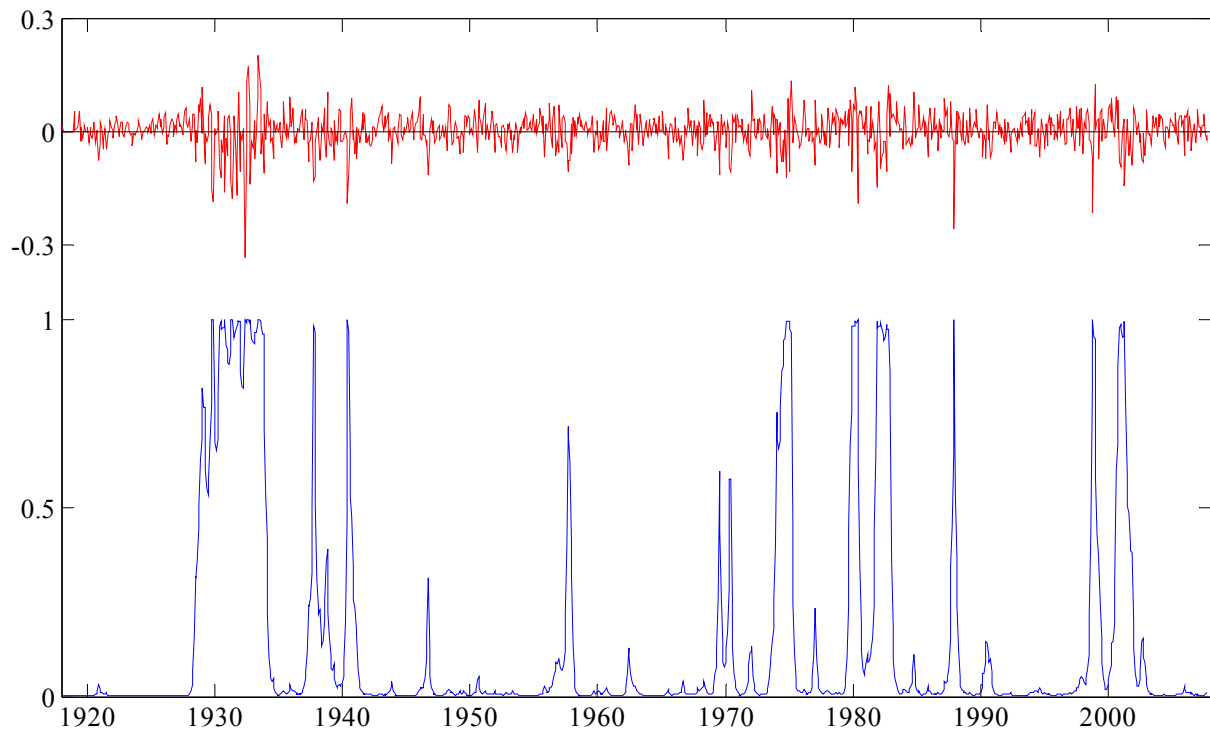
Table VI-8. The three regimes identified in Model 3-AR(1) can be organized as follows. The first regime, which has a low monthly volatility of 1.9% and a high monthly expected return of 1.2% , can be regarded as a bull market for the Canadian stock market. The second regime, which has a moderate monthly expected return of 0.7% and a medium monthly volatility of 3.6%, could be identified as a recovery regime from a bear market to a bull market. The last regime has a high monthly volatility of 8.4% but a negative monthly return of -1.7%. The relatively extreme movements of the stock market return in this regime would suggest that it is a bear market. Figure VI-15 and Figure VI-16 show the smoothed state probabilities of observing a low and high volatility regime, respectively, from Model 3- AR(1) under a 3-regime specification. One may notice the great resemblance between Figure VI-16 and Figure VI-14. This coherence of the probabilities between the two reveals the fact that recurrent shifts in regimes affect the systematic risk component of the stock market. These kinds of shifts, albeit not frequent, capture the stock return's leptokurtosis in its distribution.

Figure VI-15: Canadian stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)



The correlations between the smoothed state probabilities of observing low and high volatility regimes, across different models under the 3-regime framework, are shown in Panel B and C in Table VI-9. A significant high correlation exist across different models suggests that a 2-regime specification would be enough to incorporate the recurrent shifts in the volatility and expected return of the Canadian stock returns. In comparison to the Australian stock market returns, the economic significance of adding an additional regime in the Canadian stock market returns is relatively small.

Figure VI-16: Canadian stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)



VI.5.3 Sweden

Table VI-10 reports the parameter estimates of models 1, 2, 3 and 4 for Sweden stock market returns. Model 2-AR(1) is selected by all three information criteria as the best performing model under a 2-regime specification. Although Model 3 and 4 are more flexible models in terms of capturing the regime dependent expected mean of the stock returns, the insignificant μ_2 highlights the diminished gain of allowing a regime dependent expected mean. The Ljung-Box statistics shown in Table VI-11, however, tells that none of the models specified with 2 regimes is able to capture the autocorrelation and heteroscedasticity in the residuals. This raises some scepticism that a 2-regime specification may not be able to fully capture the heteroscedastic feature of the Sweden stock market return.

Table VI-10: Estimation result (Sweden)

$k_{AUS} = 2$	AR(0)			AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$	0.00488 (0.00137)	0.00646 (0.00113)	0.00770 (0.00128)	0.00476 (0.00164)	0.00599 (0.00130)	0.00668 (0.00152)	0.00700 (0.00147)
μ_2			-0.00083 (0.00378)			0.00037 (0.00437)	0.00018 (0.00454)
σ_1	0.04794 (0.00097)	0.03054 (0.00118)	0.03017 (0.00112)	0.04708 (0.00095)	0.03029 (0.00109)	0.03021 (0.00124)	0.03000 (0.00114)

σ_2		0.07239 (0.00360)	0.07110(0.0 0336)		0.07098 (0.00358)	0.06907 (0.00379)	0.06997 (0.00353)
p_{11}		0.97840 (0.00678)	0.97763(0.0 0701)		0.98008 (0.00646)	0.98981 (0.00518)	0.97961 (0.00661)
p_{22}		0.95410 (0.01583)	0.95518 (0.01501)		0.95541 (0.01591)	0.96683 (0.01461)	0.95674 (0.01560)
α_1				0.17847 (0.02811)	0.16532 (0.02983)	0.16364 (0.03025)	0.14080 (0.04098)
α_2							0.18811 (0.05040)
Log L	1976.564	2128.737	2130.731	1997.122	2146.545	2145.574	2147.643
AIC	-3.23434	-3.47868	-3.48031	-3.26905	-3.50909	-3.50586	-3.50761
BIC	-3.22597	-3.45777	-3.45522	-3.25649	-3.48398	-3.47656	-3.47415
HQ	-3.22155	-3.47081	-3.47087	-3.26433	-3.49964	-3.49483	-3.49501
$k_{AUS} = 3$	AR(0)			AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$		0.00620 (0.00104)	0.00732 (0.00152)		0.00434 (0.00124)	0.00698 (0.00161)	0.00700 (0.00161)
μ_2			0.00599 (0.00244)			0.00587 (0.00260)	0.00578 (0.00257)
μ_3			-0.00560 (0.00818)			-0.00454 (0.00853)	-0.00432 (0.00903)
σ_1		0.02393 (0.00171)	0.02407 (0.00176)		0.02565 (0.00239)	0.02390 (0.00217)	0.02393 (0.00222)
σ_2		0.04356 (0.00281)	0.04381 (0.00284)		0.04349 (0.00348)	0.04295 (0.00317)	0.04301 (0.00320)
σ_3		0.09105 (0.00742)	0.09040 (0.00730)		0.08823 (0.00754)	0.08638 (0.00702)	0.08647 (0.00705)
p_{11}		0.95936 (0.01633)	0.96052 (0.01587)		0.97740 (0.01640)	0.96323 (0.01850)	0.96344 (0.01888)
p_{12}		0.04063 (0.01632)	0.03948 (0.01587)		0.02195 (0.01632)	0.03677 (0.01850)	0.03656 (0.01887)
p_{21}		0.02921 (0.01202)	0.03123 (0.01232)		0.02065 (0.01285)	0.03130 (0.01276)	0.03119 (0.01294)
p_{22}		0.94482 (0.01574)	0.94354 (0.01613)		0.94769 (0.01667)	0.94289 (0.01629)	0.94317 (0.01632)
p_{31}		1.8E-11 (1.9E-08)	2.8E-09 (2.9E-06)		5.0E-07 (0.00008)	5.8E-15 (3.0E-09)	6.8E-19 (2.1E-13)
p_{32}		0.08324 (0.03376)	0.08629 (0.03462)		0.09412 (0.04040)	0.08025 (0.03286)	0.08055 (0.03315)
α_1					0.15621 (0.02968)	0.15145 (0.02975)	0.14444 (0.05453)
α_2							0.13173 (0.04905)
α_3							0.21418 (0.07762)
Log L		2154.189	2155.721		2168.547	2171.200	2171.590
AIC		-3.51219	-3.51142		-3.53696	-3.53803	-3.53539
BIC		-3.47036	-3.46122		-3.49092	-3.48362	-3.47261
HQ		-3.49644	-3.49253		-3.51963	-3.51755	-3.51176

Note: the standard errors are in parentheses. The standard errors are numerically calculated using the variance-covariance matrix of the estimators (see e.g. Hamilton (1994, p.389) for details).

Table VI-11: Ljung-Box Q statistics for autocorrelation effect and Q² statistics for ARCH effect in the residuals (Sweden)

	2-regime AR(0)		2-regime AR(1)		3-regime AR(0)		3-regime AR(1)	
	Q(12)	Q ² (12)	Q(12)	Q ² (12)	Q(12)	Q ² (12)	Q(12)	Q ² (12)
Model 2	96.581	25.436	44.982	33.048	92.714	11.962	42.619	18.470
	[0.000]	[0.013]	[0.000]	[0.001]	[0.000]	[0.449]	[0.000]	[0.102]
Model 3	85.782	25.485	44.267	33.615	85.283	12.821	39.193	18.250
	[0.000]	[0.013]	[0.000]	[0.001]	[0.000]	[0.382]	[0.000]	[0.108]
Model 4			41.291	30.820			39.458	18.348
			[0.000]	[0.002]			[0.000]	[0.106]

Note: This Ljung-Box Q statistics is a modified portmanteau test of the original one, which allows us to test the ARCH effect of the residuals. Q(12) refers to the Ljung-Box test for no serial correlation up to 12 lags. Q²(12) refers to the Ljung-Box test for no ARCH effect up to 12 lags. The p-values are shown in square brackets. A smaller p-value indicates a rejection of the null hypothesis of no serial correlation or no ARCH effect.

Now, consider the 3-regime model specification for Sweden stock market. While BIC and HQ information criteria select Model 2-AR(1) as the best in-sample performing model, AIC prefers Model 3-AR(1). The difference between Model 2-AR(1) and Model 3-AR(1) is the inclusion of μ_2 and μ_3 in the later model to capture the multi-regime dependent expected mean of the stock return. Although the expected return in second regime, μ_2 , is highly significant, the negative expected return regime with a negative μ_3 is insignificant. The evidence shown from Ljung-Box statistics in Table VI-11 suggests that Model 3-AR(1) is better than Model 2-AR(1) in capturing the heteroscedasticity effect in the residuals but, still, the residuals are serially correlated as indicated by the low p-value of the Q(12) statistics. This is probably due to the limitation of our model that we only include an autoregressive term with lag-1 in the model. Since the monthly data used in our analysis present a high serial correlation in the residuals, it would be interesting to see if the inclusion of more autoregressive terms would mitigate this problem. Or, alternatively, we may use lower frequency sampled data in analysis (e.g. quarterly or annually sampled data), which we shall leave it for future research.

The three regimes identified in Model 3 AR(1) can be organized as follows. The first regime, which has a low monthly volatility of 2.3% and a high monthly expected return of 0.58% can be regarded as a bull market for the Sweden stock market. The second regime, which has a monthly expected return (of 0.58%) less than the first regime and a monthly volatility of 4.3%, suggests that it is a recovering regime from high volatility state to a low volatility state. The last regime involves extreme movements of the Sweden stock market return. It has a high monthly volatility of 8.6% but a negative monthly return of -0.45%, which can be thought as a bear market regime.

Table VI-12: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - Sweden

	Model 2 AR(0)	Model 2 AR(1)	Model 3 AR(0)	Model 3 AR(1)	Model 4 AR(1)
2-regime (high volatility regime)					
Model 2 AR(0)	1	0.9862037	0.9940521	0.9830426	0.9944342
Model 2 AR(1)		1	0.971925	0.9911119	0.9939075
Model 3 AR(0)			1	0.97617	0.9902616
Model 3 AR(1)				1	0.9937209
Model 4 AR(1)					1
3-regime (low volatility regime)					
Model 2 AR(0)	1	0.9637577	0.9984923	0.9903814	0.9902894
Model 2 AR(1)		1	0.9639692	0.9779206	0.9783563
Model 3 AR(0)			1	0.9908679	0.9909502
Model 3 AR(1)				1	0.9999909
Model 4 AR(1)					1
3-regime (high volatility regime)					
Model 2 AR(0)	1	0.9932026	0.9981172	0.9919514	0.9955006
Model 2 AR(1)		1	0.9893094	0.9961146	0.9962558
Model 3 AR(0)			1	0.9916795	0.9950454
Model 3 AR(1)				1	0.9994121
Model 4 AR(1)					1

Figure VI-17 and Figure VI-18 show the smoothed state probabilities of observing a low and high volatility regime, respectively, from Model 3-AR(1) under a 3-regime specification. Similar to the cases in Australian and Canadian stock markets, the high volatility regime under the 3-regime framework captures the extreme movements of the Sweden stock market which are short-lived. This helps us to capture the stock return's leptokurtosis in its distribution. In addition, as most of the extreme movements in the stock market are downward drops, Markov switching models with 3 regimes would better capture the negative skewness in the stock returns than their 2-regime counterparts. The correlations between the smoothed state probabilities of observing low and high volatility regimes, across different models under the 3-regime framework, are shown in Panel B and C in Table VI-12. A significant high correlation exists across different models.

Figure VI-17: Sweden stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)

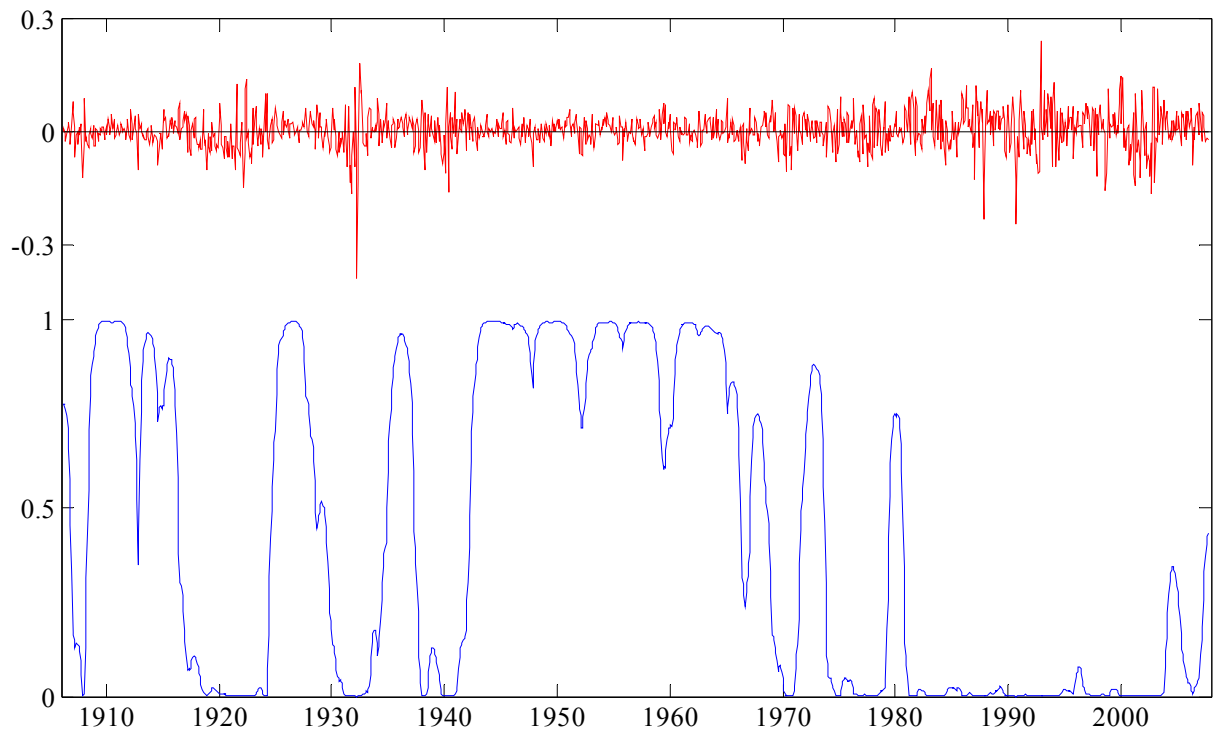
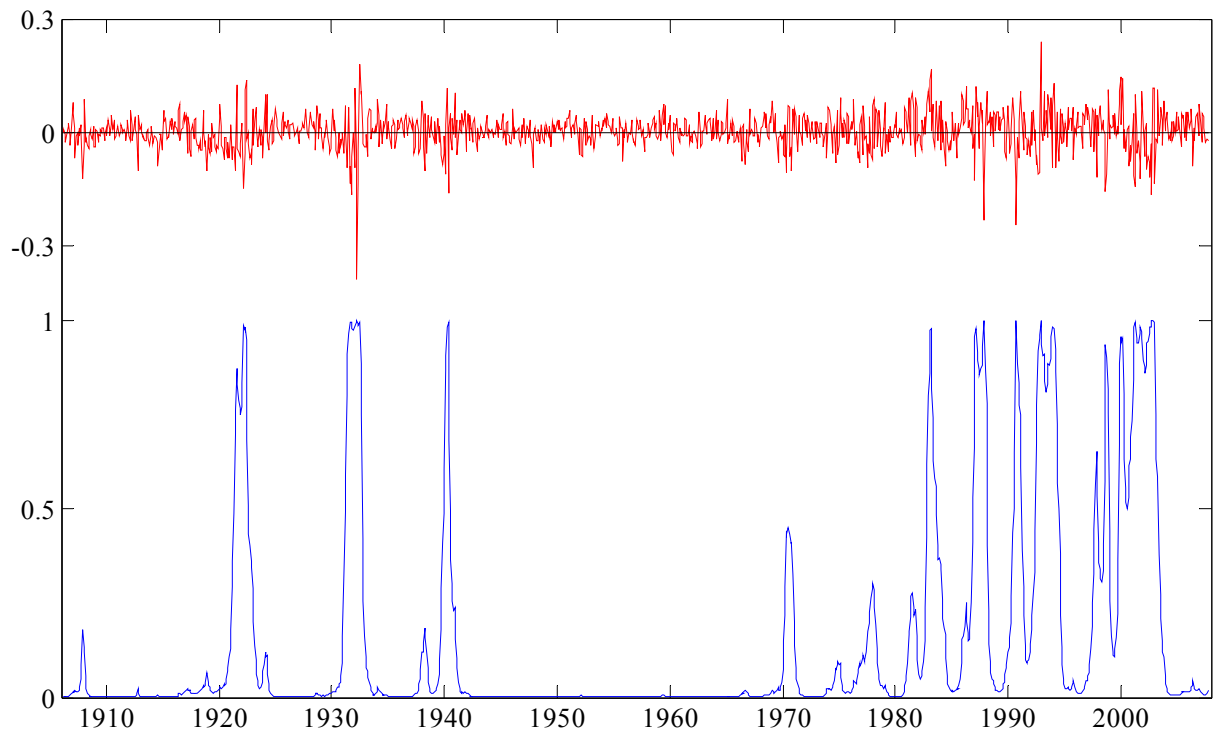


Figure VI-18: Sweden stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)



VI.5.4 Switzerland

Table VI-13 reports the parameter estimates on Switzerland stock market returns. While Model 3-AR(1) is selected as the best performing model under a 2-regime specification by the AIC and HQ information criterion, BIC prefers Model 2-AR(1). In addition, the insignificant μ_2 in Model 3-AR(1) suggest that the expected return of the Switzerland stock market may not be affected by shifting regimes in the volatility. The Ljung-Box statistics, in Table VI-14, suggests Model 3-AR(1) is more capable than Model 2-AR(1) to tackle the heteroscedasticity problem in the residuals. However, substantive autocorrelation in the residuals still remains.

Table VI-13: Estimation result (Switzerland)

$k_{AUS} = 2$	AR(0)			AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$	0.00327 (0.00131)	0.00539 (0.00105)	0.00673 (0.00116)	0.00328 (0.00153)	0.00540 (0.00119)	0.00664 (0.00129)	0.00663 (0.00128)
μ_2			-0.00482 (0.00419)			-0.00485 (0.00463)	-0.00473 (0.00471)
σ_1	0.04359 (0.00093)	0.02770 (0.00112)	0.02775 (0.00109)	0.04311 (0.00092)	0.02756 (0.00109)	0.02755 (0.00109)	0.02754 (0.00109)
σ_2		0.06720 (0.00404)	0.06671 (0.00385)		0.06647 (0.00385)	0.06619 (0.00379)	0.06609 (0.00379)
p_{11}		0.96282 (0.01166)	0.96242 (0.01173)		0.96240 (0.01179)	0.96135 (0.01213)	0.96147 (0.01212)
p_{22}		0.91326 (0.03420)	0.91076 (0.03447)		0.91157 (0.03263)	0.90764 (0.03437)	0.90840 (0.03442)
α_1				0.15026 (0.02982)	0.13361 (0.03015)	0.12745 (0.02999)	0.11769 (0.03731)
α_2							0.15013 (0.05896)
Log L	1885.383	2007.827	2011.411	1895.720	2014.772	2017.637	2017.736
AIC	-3.42433	-3.64150	-3.64620	-3.44444	-3.65564	-3.65903	-3.65739
BIC	-3.41524	-3.61876	-3.61891	-3.43079	-3.62833	-3.62717	-3.62098
HQ	-3.41028	-3.63290	-3.63588	-3.43928	-3.64530	-3.64698	-3.64361
$k_{AUS} = 3$	AR(0)			AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$		0.00530 (0.00105)	0.00650 (0.00119)		0.00523 (0.00118)	0.00292 (0.00219)	0.00264 (0.00259)
μ_2			0.00208 (0.00104)			0.00725 (0.00170)	0.00766 (0.00166)
μ_3			-0.04730 (0.02848)			-0.00703 (0.00633)	-0.01089 (0.00749)
σ_1		0.02637 (0.00141)	0.02629 (0.00103)			0.01491 (0.00257)	0.01630 (0.00166)
σ_2		0.05142 (0.02278)	0.05110 (0.00282)		0.05066 (0.01026)	0.03263 (0.00371)	0.03229 (0.00262)
σ_3		0.13541 (0.11566)	0.12040 (0.01395)		0.13881 (0.07000)	0.07155 (0.00679)	0.07437 (0.00612)
p_{11}		0.97163 (0.02108)	0.96881 (0.01194)		0.97114 (0.01463)	0.87142 (0.08730)	0.94480 (0.02651)
p_{12}		0.01975 (0.01738)	0.02279 (0.01198)		0.02057 (0.01613)	0.12846 (0.08735)	0.05520 (0.02651)

p_{21}	0.03693 (0.05812)	0.04615 (0.01901)	0.04593 (0.02716)	0.03813 (0.02515)	0.01347 (0.00732)
p_{22}	0.92663 (0.07713)	0.92535 (0.01792)	0.91209 (0.09971)	0.92914 (0.02722)	0.93641 (0.01803)
p_{31}	0.18133 (0.55678)	0.11079 (0.08707)	0.00876 (0.07637)	0.00119 (0.00354)	2.0E-09 (1.4E-06)
p_{32}	0.63723 (1.12721)	0.51227 (0.26178)	0.98959 (0.08132)	0.09787 (0.05002)	0.16680 (0.06830)
α_1			0.13585 (0.02950)	0.09036 (0.03093)	0.28330 (0.09420)
α_2					0.09868 (0.04031)
α_3					0.15220 (0.07786)
Log L	2021.414	2025.213	2028.883	2030.827	2036.250
AIC	-3.65712	-3.66039	-3.67222	-3.67211	-3.67835
BIC	-3.61163	-3.60581	-3.62215	-3.61294	-3.61007
HQ	-3.63991	-3.63974	-3.65327	-3.64973	-3.65252

Note: the standard errors are in parentheses. The standard errors are numerically calculated using the variance-covariance matrix of the estimators (see e.g. Hamilton (1994, p.389) for details).

Table VI-14: Ljung-Box Q statistics for autocorrelation effect and Q^2 statistics for ARCH effect in the residuals (Switzerland)

	2-regime AR(0)		2-regime AR(1)		3-regime AR(0)		3-regime AR(1)	
	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$
Model 2	53.435 [0.000]	9.761 [0.637]	30.465 [0.002]	9.913 [0.624]	57.264 [0.000]	10.123 [0.605]	32.325 [0.001]	8.520 [0.743]
Model 3	46.862 [0.000]	8.347 [0.757]	28.182 [0.005]	8.820 [0.718]	49.984 [0.000]	9.962 [0.619]	30.070 [0.003]	4.681 [0.968]
Model 4			28.182 [0.005]	9.042 [0.699]			28.099 [0.005]	7.134 [0.849]

Note: This Ljung-Box Q statistics is a modified portmanteau test of the original one, which allows us to test the ARCH effect of the residuals. Q(12) refers to the Ljung-Box test for no serial correlation up to 12 lags. $Q^2(12)$ refers to the Ljung-Box test for no ARCH effect up to 12 lags. The p-values are shown in square brackets. A smaller p-value indicates a rejection of the null hypothesis of no serial correlation or no ARCH effect.

Similar to Australian, Canadian and Sweden stock markets, the high volatility regime identified in the Switzerland stock market is also related to a low expected return. The monthly expected return in a high volatility regime is -0.485%, while in the low volatility regime is 0.664%. Figure VI-19 plots the smoothed state probabilities of observing a high volatility regime filtered from Model 3-AR(1). There is a clear matching between the probability of observing a high volatility regime and many well-known historical events as we find in other countries. In addition, the high volatility regime in Switzerland stock market is more frequently visited than it is in the aforementioned countries.

Figure VI-19: Switzerland stock market return series and Model 3-AR(1)'s smoothed posterior probability of observing a high volatility regime (2-regime)

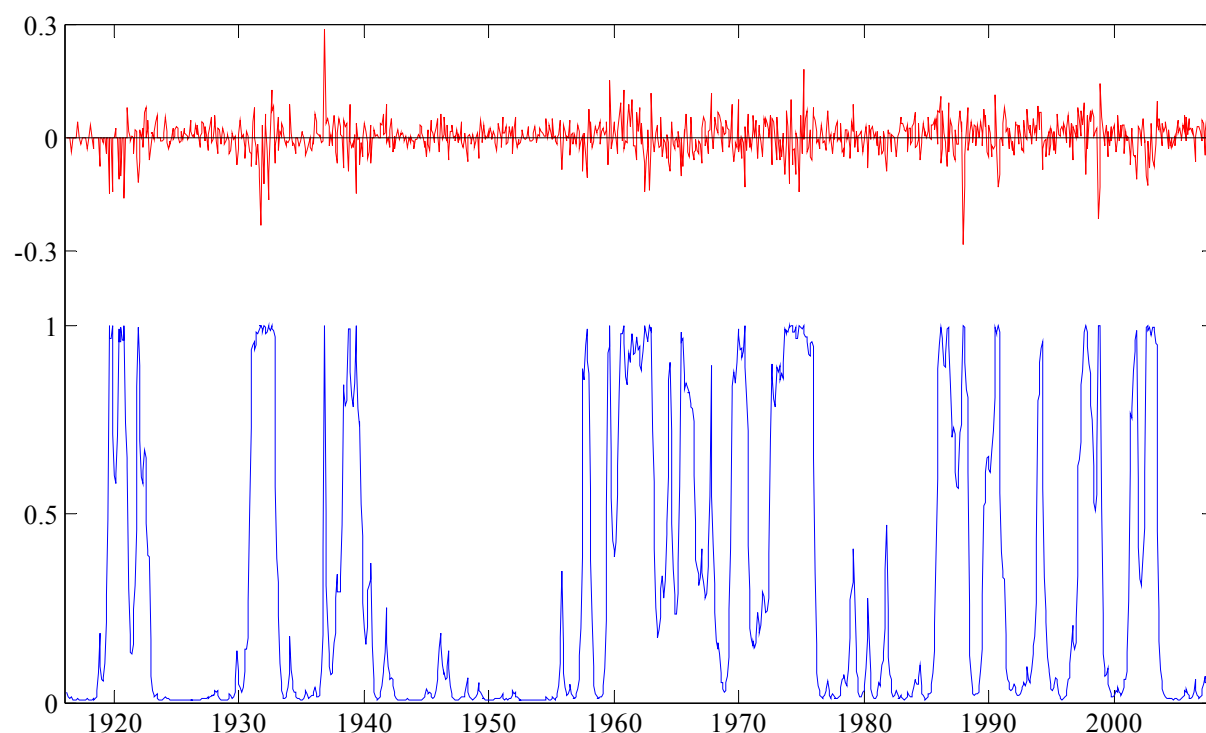


Table VI-15: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - Switzerland

	Model 2 AR(0)	Model 2 AR(1)	Model 3 AR(0)	Model 3 AR(1)	Model 4 AR(1)
2-regime (high volatility regime)					
Model 2 AR(0)	1	0.99510862	0.99519807	0.9919321	0.9931875
Model 2 AR(1)		1	0.99094513	0.9967061	0.9968184
Model 3 AR(0)			1	0.9954653	0.9964742
Model 3 AR(1)				1	0.9999175
Model 4 AR(1)					1
3-regime (low volatility regime)					
Model 2 AR(0)	1	0.99619873	0.99656359	0.5297908	0.4813523
Model 2 AR(1)		1	0.99424949	0.5392236	0.4920527
Model 3 AR(0)			1	0.5299908	0.4818896
Model 3 AR(1)				1	0.9424528
Model 4 AR(1)					1
3-regime (high volatility regime)					
Model 2 AR(0)	1	0.90738694	0.9695129	0.4558541	0.5099831
Model 2 AR(1)		1	0.8197015	0.4199568	0.4652378
Model 3 AR(0)			1	0.5042298	0.5622091
Model 3 AR(1)				1	0.9791854
Model 4 AR(1)					1

Now, consider the models with a 3-regime specification. While AIC information criterion select Model 4-AR(1) as the best fitting model, BIC and HQ prefer Model 2-AR(1). The difference between Model 4-AR(1) and Model 2-AR(1) is not only the inclusion of μ_2 and μ_3 in the former model to capture the regime dependent expected mean of the stock return, but also the regime dependent autoregressive term coefficients α_1 and α_2 . The idea of including the regime dependent α_1 and α_2 in Model 4-AR(1) is due to the consideration of detecting the persistence of the stock returns within each distinguished volatility regimes. The statistically significant α_1 and α_2 suggest that the predictability of the Switzerland stock market return still exists even the regime shifts in the expected return are taken into account. The evidence shown from Ljung-Box statistics in Table VI-14 suggests that Model 4-AR(1) can significantly capture the heteroscedasticity effect in the residuals, but not the autocorrelation in the residuals at a 12-lag horizon. Similar to the Sweden case, this is probably due to the limitation of our model that we only include an autoregressive term with lag-1 in the model. Since the monthly data used in our analysis present a high serial correlation in the residuals, it would be interesting to see if the inclusion of more autoregressive terms would mitigate this problem. Or, alternatively, we may use lower frequency sampled data in analysis (e.g. quarterly or annually sampled data), which we shall leave it for future research.

The three regimes identified in Model 4-AR(1) can be organized as follows. The first regime has a low monthly volatility of 1.63% and a moderate monthly expected return of 0.264%. The second regime has a monthly expected return of 0.76%, which is the highest amongst the three regimes, and a moderate monthly volatility of 3.2%. The last regime has a high monthly volatility of 7.4% but with a negative monthly return of -1.09%. The relatively extreme movements of the stock market return in this regime would suggest it as a bear market. In different to the three regimes identified in other countries, the regimes in Switzerland stock market seem non-standard - the duration of the first and second regimes combined (which have a relatively low volatility but positive expected return) account for 33.84 months. Visually, the smoothed state probability of observing a low volatility regime, as plotted in Figure VI-20, further assists this point. However, one thing should be kept in mind is that the correlations between Model 4-AR(1)'s filtered probability of observing a low/high volatility regime and those probabilities filtered from other models are quite low.

Figure VI-20: Switzerland stock market return series and Model 4-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)

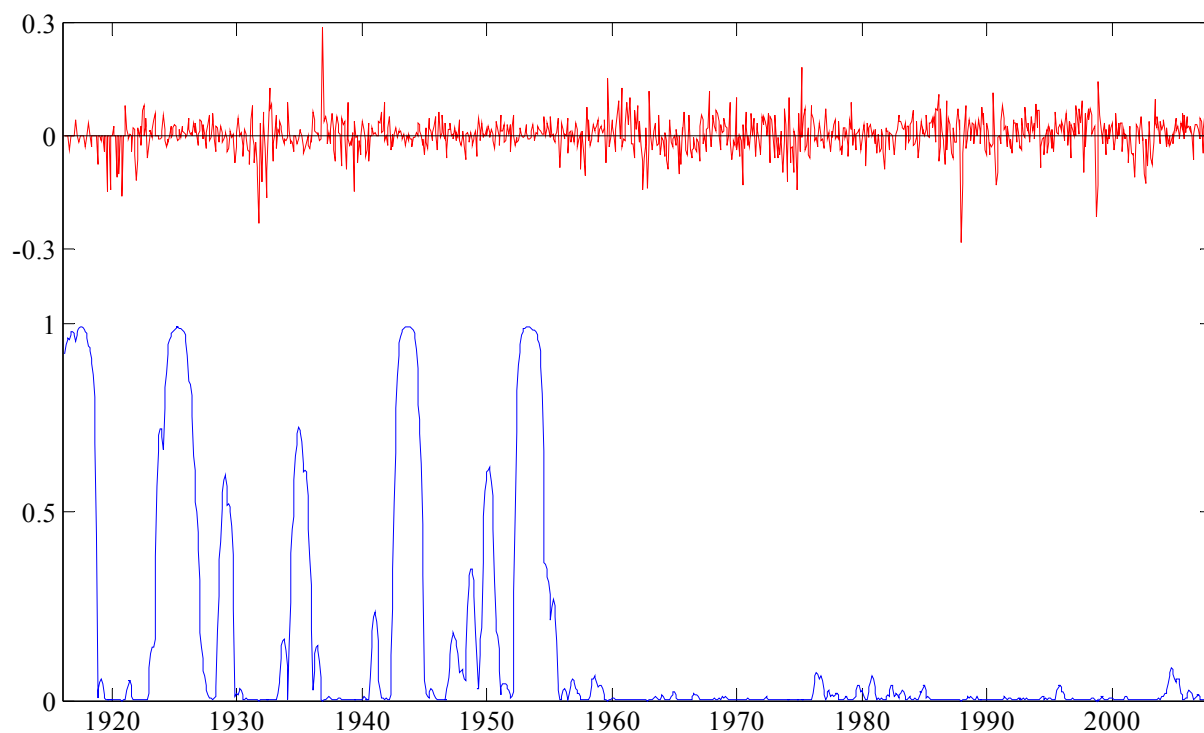
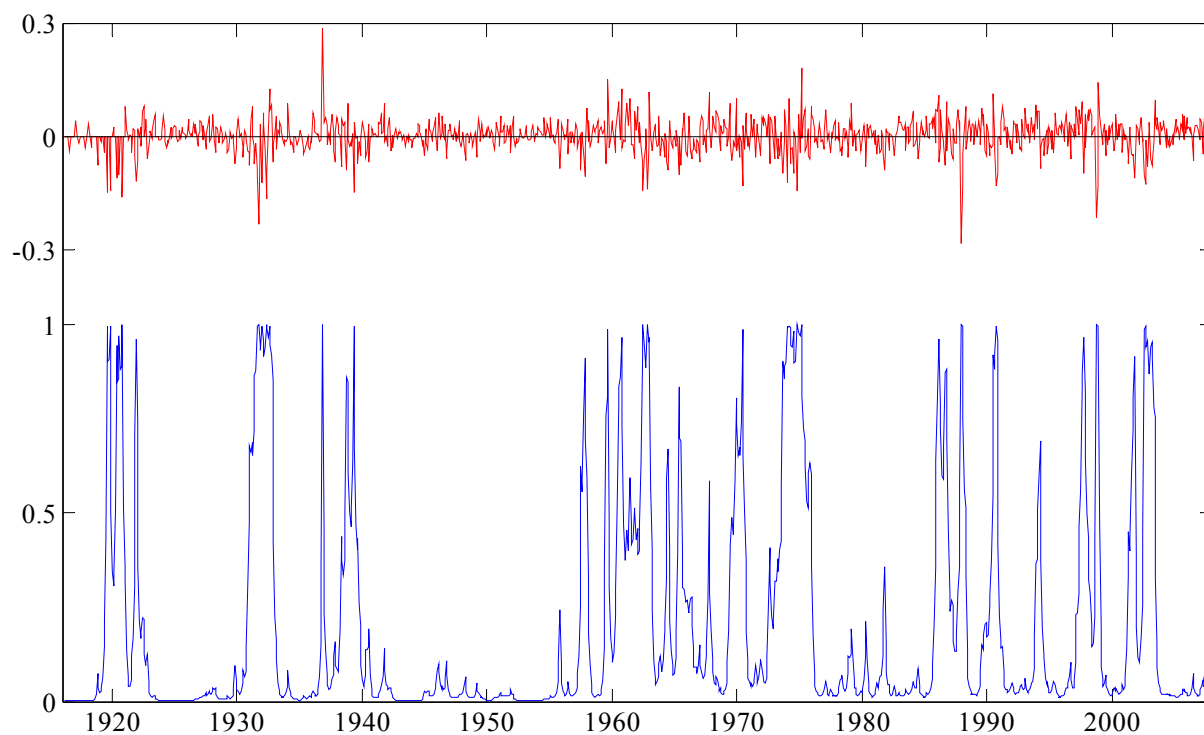


Figure VI-21: Switzerland stock market return series and Model 4-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)



VI.5.5 UK

Table VI-16 reports the parameter estimation results for UK stock market returns. Looking at panels A1 and A2, Model 2-AR(1) is selected by all three information criteria as the best performing model under a 2-regime specification. Although Model 3 and 4 are more flexible models in terms of capturing the regime dependent expected mean of the stock returns, the insignificant μ_2 highlights the diminished gain of allowing a regime dependent expected mean. The Ljung-Box statistics, in Table VI-17, show that there are still significant levels of autocorrelation and heteroscedasticity remained in the residuals from 2-regime models, which suggests a 2-regime model would not be sufficient to capture the heteroscedastic feature in UK stock return data.

Table VI-16: Estimation result (UK)

$k_{AUS} = 2$	AR(0)			AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$	0.00210 (0.00075)	0.00216 (0.00044)	0.00218 (0.00046)	0.00210 (0.00090)	0.00219 (0.00053)	0.00218 (0.00055)	0.00222 (0.00057)
μ_2			0.00191 (0.00194)			0.00441 (0.00237)	0.00188 (0.00226)
σ_1	0.03760 (0.00053)	0.01648 (0.00045)	0.01648 (0.00045)	0.03700 (0.00050)	0.01625 (0.00047)	0.01627 (0.00048)	0.01622 (0.00048)
σ_2		0.05788 (0.00157)	0.05788 (0.00157)		0.05696 (0.00156)	0.05694 (0.00156)	0.05683 (0.00156)
p_{11}		0.97810 (0.00489)	0.97810 (0.00489)		0.97987 (0.00473)	0.98039 (0.00467)	0.97984 (0.00474)
p_{22}		0.96204 (0.00846)	0.96202 (0.00847)		0.96525 (0.00800)	0.96571 (0.00792)	0.96541 (0.00796)
α_1				0.17040 (0.01970)	0.19790 (0.02085)	0.19844 (0.02083)	0.22479 (0.02799)
α_2							0.16142 (0.03281)
Log L	4642.783	5365.751	5365.760	4677.275	5407.869	5407.325	5408.906
AIC	-3.72454	-4.30237	-4.30157	-3.75293	-4.33711	-4.33587	-4.33634
BIC	-3.71987	-4.29069	-4.28756	-3.74592	-4.32309	-4.31951	-4.31764
HQ	-3.71790	-4.29813	-4.29648	-3.75039	-4.33202	-4.32993	-4.32955
$k_{AUS} = 3$	AR(0)			AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$		0.00128 (0.00042)	0.00153 (0.00047)		0.00175 (0.00049)	0.00145 (0.00053)	0.00140 (0.00058)
μ_2			0.00373 (0.00121)			0.00370 (0.00142)	0.00383 (0.00130)
μ_3			-0.00681 (0.01855)			-0.01558 (0.02008)	-0.01861 (0.02193)
σ_1		0.01334 (0.00050)	0.01433 (0.00042)		0.01397 (0.00039)	0.01393 (0.00039)	0.01389 (0.00039)
σ_2		0.02929 (0.00287)	0.03926 (0.00121)		0.03896 (0.00120)	0.03881 (0.00119)	0.03871 (0.00113)
σ_3		0.06894 (0.00545)	0.14552 (0.01540)		0.14058 (0.01498)	0.13929 (0.01481)	0.13960 (0.01453)
p_{11}		0.98347 (0.00512)	0.97929 (0.00512)		0.98006 (0.00498)	0.98083 (0.00488)	0.98116 (0.00486)

p_{12}	0.01653 (0.00512)	0.02071 (0.00512)	0.01994 (0.00498)	0.01828 (0.00485)	0.01795 (0.00486)
p_{21}	0.01735 (0.00559)	0.02195 (0.00603)	0.02099 (0.00576)	0.01991 (0.00553)	0.01937 (0.00499)
p_{22}	0.95055 (0.01349)	0.96970 (0.00692)	0.96961 (0.00694)	0.97190 (0.00660)	0.97286 (0.00642)
p_{31}	0.00053 (0.00162)	0.00012 (0.00150)	3.1E-09 (1.9E-06)	0.00005 (0.00091)	5.5E-41 (2.6E-34)
p_{32}	0.07715 (0.04761)	0.14939 (0.05275)	0.15990 (0.05564)	0.15463 (0.05349)	0.15053 (0.05196)
α_1			0.17812 (0.02077)	0.17639 (0.02063)	0.25935 (0.03090)
α_2					0.09656 (0.02955)
α_3					0.24594 (0.11453)
Log L	5479.421	5524.041	5556.410	5559.122	5566.138
AIC	-4.38958	-4.42379	-4.45236	-4.45293	-4.45696
BIC	-4.36623	-4.39576	-4.42665	-4.42255	-4.42191
HQ	-4.38110	-4.41361	-4.44302	-4.44190	-4.44423

Note: the standard errors are in parentheses. The standard errors are numerically calculated using the variance-covariance matrix of the estimators (see e.g. Hamilton (1994, p.389) for details).

Table VI-17: Ljung-Box Q statistics for autocorrelation effect and Q^2 statistics for ARCH effect on the residuals (UK)

	2-regime AR(0)		2-regime AR(1)		3-regime AR(0)		3-regime AR(1)	
	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$
Model 2	120.352 [0.000]	255.320 [0.000]	25.756 [0.012]	191.551 [0.000]	128.557 [0.000]	75.533 [0.000]	32.887 [0.001]	3.863 [0.986]
Model 3	120.419 [0.000]	255.444 [0.000]	25.627 [0.012]	189.423 [0.000]	97.368 [0.000]	5.361 [0.945]	32.132 [0.001]	4.743 [0.966]
Model 4			25.156 [0.014]	199.742 [0.000]			30.667 [0.002]	4.165 [0.980]

Note: This Ljung-Box Q statistics is a modified portmanteau test of the original one, which allows us to test the ARCH effect of the residuals. Q(12) refers to the Ljung-Box test for no serial correlation up to 12 lags. $Q^2(12)$ refers to the Ljung-Box test for no ARCH effect up to 12 lags. The p-values are shown in square brackets. A smaller p-value indicates a rejection of the null hypothesis of no serial correlation or no ARCH effect.

Now, consider the estimation results under a 3-regime specification in panel B1 and B2 of Table VI-16. While BIC information criterion select Model 2-AR(1) as the best fitting model, AIC and HQ prefer Model 4-AR(1). The difference between Model 4-AR(1) and Model 2-AR(1) is not only the inclusion of μ_2 and μ_3 in the former model to capture the regime dependent expected mean of the stock return, but also the regime dependent autoregressive term coefficients α_1 and α_2 . Although the negative expected stock return μ_3 is not statistically significant, the statistically significant α_1 and α_2 suggest that the predictability of the UK stock market return still exists even the regime shifts in the expected return are taken into account. The evidence shown from Ljung-Box statistics in Table VI-17 suggests that Model 4-AR(1) can significantly capture the heteroscedasticity effect in the

residuals, but far from being able to tackle the autocorrelation problem in the residuals at a 12-lag horizon as seen in other models. This is still the problem we found in the Sweden and Switzerland cases, which is probably due to the limitation of our model that we only include an autoregressive term with lag-1 in the model. Since the monthly data used in our analysis present a high serial correlation in the residuals, it would be interesting to see if the inclusion of more autoregressive terms would mitigate this problem. Or, alternatively, we may use lower frequency sampled data in analysis (e.g. quarterly or annually sampled data), which we shall leave it for future research.

The three regimes identified in Model 4-AR(1) can be organized as follows. The first regime has a low monthly volatility of 1.389% and a moderate monthly expected return of 0.14%. The second regime has a monthly expected return of 0.383%, which is the highest amongst the three regimes, and a moderate monthly volatility of 3.871%. The last regime has a high monthly volatility of 13.96% but with a negative monthly return of -1.861%. The relatively extreme movements of the stock market return in this regime would suggest it as a bear market. In different to the three regimes identified for the previous countries, the regimes in the UK stock market seem non-standard - the duration of the low volatility and high return regime is 53 months. Visually, the smoothed state probability of observing a low and high volatility regime, as plotted in Figure VI-22 and Figure VI-23, respectively, further assists this point. There is a clear matching between the probability of observing a high volatility regime and many well-known historical events, e.g. the 1825 banking crisis, the 1929-1933 Great Depression, the WWII, the 1973 first oil crisis and 1973-75 second banking crisis, the 1979 second oil crisis and the 1987 stock market crash. The correlations between Model 4-AR(1)'s smoothed probability of observing a low/high volatility regime and those probabilities from other models are shown in Table VI-18, which are quite high.

Table VI-18: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - UK

	Model 2 AR(0)	Model 2 AR(1)	Model 3 AR(0)	Model 3 AR(1)	Model 4 AR(1)
2-regime (high volatility regime)					
Model 2 AR(0)	1	0.9911589	0.9999988	0.99103223	0.98811988
Model 2 AR(1)		1	0.9911919	0.99996996	0.99971825
Model 3 AR(0)			1	0.99105797	0.98815522
Model 3 AR(1)				1	0.99967361
Model 4 AR(1)					1
3-regime (low volatility regime)					
Model 2 AR(0)	1	0.9446511	0.9493918	0.94683583	0.94339359

Model 2 AR(1)	1	0.9939553	0.9986976	0.99593837
Model 3 AR(0)		1	0.99205641	0.98386513
Model 3 AR(1)			1	0.99795365
Model 4 AR(1)				1
3-regime (high volatility regime)				
Model 2 AR(0)	1	0.4921993	0.4756945	0.49230175
Model 2 AR(1)		1	0.9916561	0.99888707
Model 3 AR(0)			1	0.99290806
Model 3 AR(1)				1
Model 4 AR(1)				1

Figure VI-22: The UK stock market return series and Model 4-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)

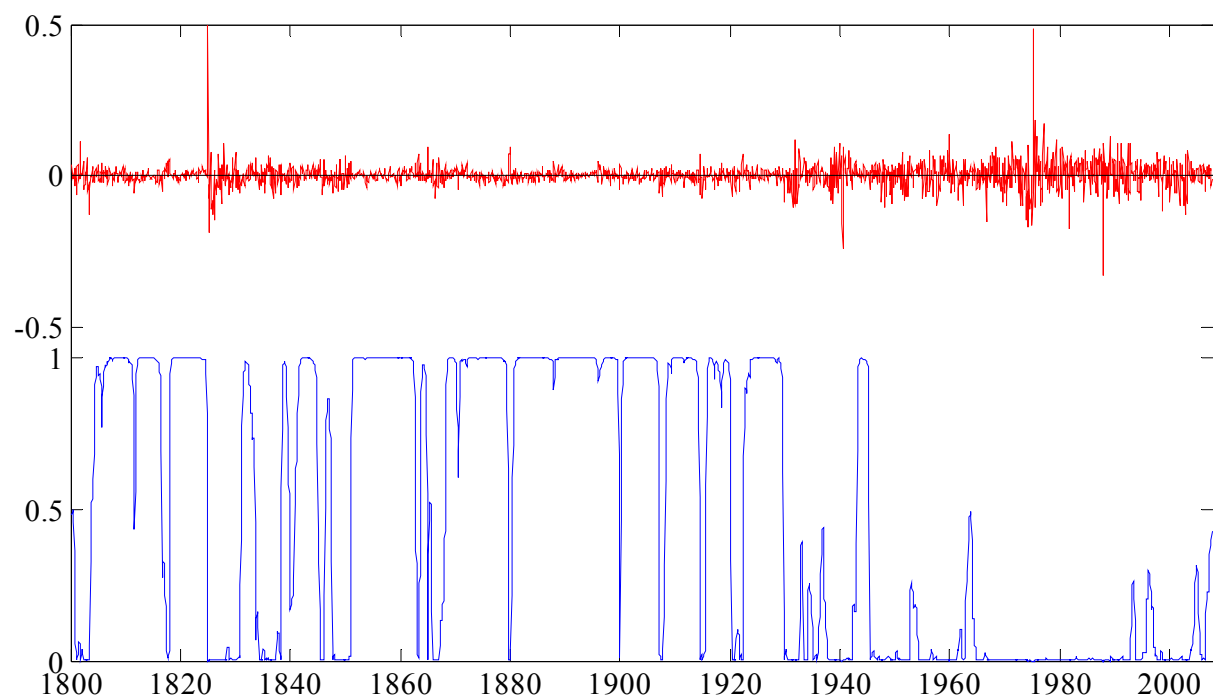
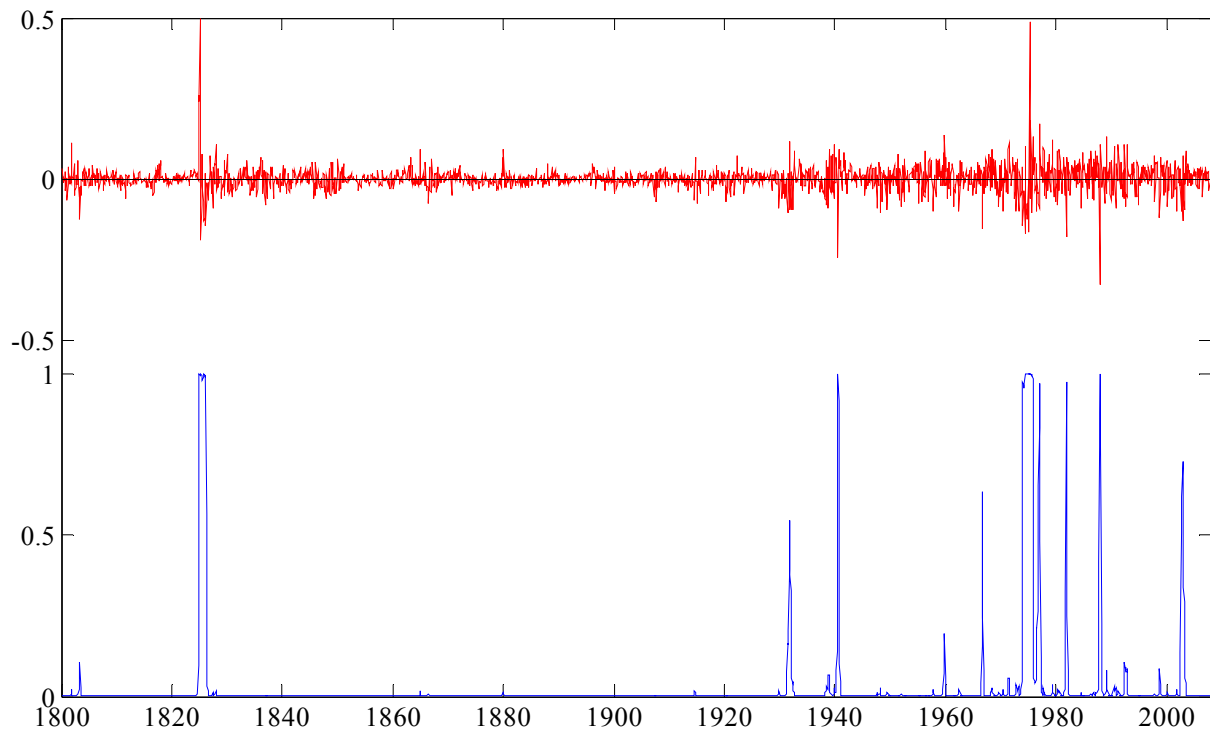


Figure VI-23: The UK stock market return series and Model 4-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)



VI.5.6 US

Table VI-19 reports the parameter estimates of models 1, 2, 3 and 4 for the US stock market returns. While Model 3-AR(1) is selected by the AIC and HQ information criteria as the best in-sample performing model under a 2-regime specification, BIC prefers Model 2-AR(1). The Ljung-Box statistics, in Table VI-20, suggests Model 3-AR(1) is more capable than Model 2-AR(1) to tackle the autocorrelation effect in the residuals. However, the heteroskedasticity in the residuals still remains.

Table VI-19: Estimation result (US)

$k_{AUS} = 2$	AR(0)			AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$	0.00258 (0.00087)	0.00419 (0.00069)	0.00484 (0.00071)	0.00258 (0.00098)	0.00404 (0.00078)	0.00463 (0.00080)	0.00462 (0.00081)
μ_2			-0.01049 (0.00535)			-0.00948 (0.00575)	-0.00920 (0.00540)
σ_1	0.04354 (0.00062)	0.02929 (0.00090)	0.02951 (0.00081)	0.04323 (0.00061)	0.02863 (0.00091)	0.02890 (0.00085)	0.02881 (0.00086)
σ_2		0.08662 (0.00530)	0.08730 (0.00481)		0.08365 (0.00514)	0.08528 (0.00484)	0.08460 (0.00479)
p_{11}		0.97649 (0.00579)	0.97675 (0.00559)		0.97598 (0.00566)	0.97613 (0.00567)	0.97588 (0.00570)
p_{22}		0.87268 (0.02789)	0.86493 (0.02837)		0.88238 (0.02799)	0.87203 (0.02799)	0.87351 (0.02790)

α_1				0.11986 (0.01989)	0.14385 (0.02102)	0.14012 (0.02106)	0.15461 (0.02416)
α_2							0.07837 (0.05324)
Log L	4274.384	4681.722	4686.194	4290.194	4702.192	4705.493	4706.291
AIC	-3.42888	-3.75339	-3.75617	-3.44215	-3.77053	-3.77238	-3.77221
BIC	-3.42421	-3.74171	-3.74216	-3.43514	-3.75651	-3.75602	-3.75352
HQ	-3.42223	-3.74915	-3.75109	-3.43960	-3.76544	-3.76644	-3.76543
$k_{AUS} = 3$	AR(0)			AR(1)			
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 4
$\mu_1(\mu)$		0.00265 (0.00055)	0.00175 (0.00070)		0.00236 (0.00063)	0.00167 (0.00080)	0.00167 (0.00077)
μ_2			0.00473 (0.00101)			0.00483 (0.00113)	0.00469 (0.00112)
μ_3			-0.01796 (0.01021)			-0.00812 (0.01072)	-0.01798 (0.01079)
σ_1		0.01644 (0.00062)	0.01641 (0.00062)		0.01649 (0.00062)	0.01634 (0.00063)	0.01640 (0.00062)
σ_2		0.03895 (0.00114)	0.05505 (0.00113)		0.03862 (0.00115)	0.03783 (0.00130)	0.03825 (0.00119)
σ_3		0.11873 (0.00825)	0.17133 (0.00829)		0.11777 (0.00817)	0.11111 (0.00785)	0.11478 (0.00790)
p_{11}		0.97963 (0.00731)	0.98070 (0.00711)		0.98059 (0.00731)	0.98160 (0.00757)	0.98147 (0.00708)
p_{12}		0.02011 (0.00731)	0.01930 (0.00711)		0.01857 (0.00727)	0.01711 (0.00863)	0.01853 (0.00709)
p_{21}		0.00901 (0.00325)	0.00843 (0.00313)		0.00958 (0.00345)	0.00725 (0.00364)	0.00805 (0.00316)
p_{22}		0.98450 (0.00419)	0.98405 (0.00446)		0.98482 (0.00427)	0.98325 (0.00492)	0.98470 (0.00441)
p_{31}		1.2E-08 (0.00002)	2.4E-32 (5.5E-27)		9.9E-09 (4.9E-06)	0.00705 (0.01052)	7.8E-63 (9.3E-57)
p_{32}		0.07670 (0.02993)	0.08626 (0.03412)		0.08388 (0.03063)	0.10834 (0.03767)	0.08122 (0.03276)
α_1					0.12335 (0.02079)	0.11978 (0.02094)	0.09653 (0.04593)
α_2							0.12942 (0.02543)
α_3							0.08569 (0.08507)
Log L		4856.123	4860.791		4869.219	4872.051	4874.224
AIC		-3.88934	-3.89149		-3.90062	-3.90129	-3.90142
BIC		-3.86599	-3.86346		-3.87492	-3.87091	-3.86638
HQ		-3.88086	-3.88131		-3.89128	-3.89026	-3.88870

Note: the standard errors are in parentheses. The standard errors are numerically calculated using the variance-covariance matrix of the estimators (see e.g. Hamilton (1994, p.389) for details).

Table VI-20: Ljung-Box Q statistics for autocorrelation effect and Q^2 statistics for ARCH effect on the residuals (US)

	2-regime AR0		2-regime AR1		3-regime AR0		3-regime AR1	
	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$	Q(12)	$Q^2(12)$
Model 2	59.913 [0.000]	62.718 [0.000]	20.604 [0.056]	61.677 [0.000]	65.926 [0.000]	5.720 [0.930]	24.848 [0.016]	6.385 [0.895]
Model 3	55.403 [0.000]	68.694 [0.000]	20.022 [0.067]	66.739 [0.000]	54.722 [0.000]	7.606 [0.815]	20.269 [0.062]	11.239 [0.509]
Model 4			20.186	70.025			20.274	6.907

	[0.066]	[0.000]	[0.062]	[0.864]
--	---------	---------	---------	---------

Note: This Ljung-Box Q statistics is a modified portmanteau test of the original one, which allows us to test the ARCH effect of the residuals. Q(12) refers to the Ljung-Box test for no serial correlation up to 12 lags. Q²(12) refers to the Ljung-Box test for no ARCH effect up to 12 lags. The p-values are shown in square brackets. A smaller p-value indicates a rejection of the null hypothesis of no serial correlation or no ARCH effect.

Now, consider the estimation results for models under a 3-regime specification in panel B1 and B2 of Table VI-19. While BIC and HQ information criteria select Model 2-AR(1) as the best in-sample performing model, AIC prefer Model 4-AR(1). The difference between Model 4-AR(1) and Model 2-AR(1) is not only the inclusion of μ_2 and μ_3 in the former model to capture the regime dependent expected mean of the stock return, but also the regime dependent autoregressive term coefficients α_1 and α_2 . Although the negative expected stock return μ_3 and the autoregressive term coefficient in regime 3 (α_3) are not statistically significant, the evidence shown from Ljung-Box statistics in Table VI-20 suggests that Model 4-AR(1) is better than Model 2-AR(1) in capturing the autocorrelation and heteroscedasticity effect in the residuals.

The three regimes identified in Model 4 AR(1) can be organized as follows. The first regime has a low monthly volatility of 1.64% and a moderate monthly expected return of 0.167%. The second regime has a monthly expected return of 0.469%, which is the highest amongst the three regimes, and a moderate monthly volatility of 3.825%. The last regime has a high monthly volatility of 11.478% but with a negative monthly return of -1.798%. The relatively extreme movements of the stock market return in this regime would suggest it as a bear market. Similar to the three regimes identified in UK stock market, the regimes in US stock market seem non-standard - the duration of the low volatility and high return regime is almost 54 months. The smoothed state probability of observing a low and high volatility regime, as plotted in Figure VI-24 and Figure VI-25, respectively, further assists this point. The correlations between Model 4-AR(1)'s filtered probability of observing a low/high volatility regime and those probabilities filtered from other models are shown in Table VI-21, which are quite high. There is also a clear matching between the probability of observing a high volatility regime and many well-known historical events, e.g. the financial panic of 1875, the American civil war, the 1929-1933 Great Depression, the WWII, the 1973 first oil crisis and 1973-75 second banking crisis, the 1987 stock market crash and the recent 2000 Dot-com bubble.

Table VI-21: Correlation between smoothed probability of high (low and high) volatility regime across 2-regime (3-regime) models - US

	Model 2 AR(0)	Model 2 AR(1)	Model 3 AR(0)	Model 3 AR(1)	Model 4 AR(1)
2-regime (high volatility regime)					
Model 2 AR(0)	1	0.984295074	0.9951948	0.987685098	0.98310432
Model 2 AR(1)		1	0.9735848	0.99508394	0.9958598
Model 3 AR(0)			1	0.985994994	0.98065856
Model 3 AR(1)				1	0.9994997
Model 4 AR(1)					1
3-regime (low volatility regime)					
Model 2 AR(0)	1	0.996230128	0.9992205	0.992195017	0.99579521
Model 2 AR(1)		1	0.9968145	0.997849932	0.99924506
Model 3 AR(0)			1	0.994849414	0.99762079
Model 3 AR(1)				1	0.9992396
Model 4 AR(1)					1
3-regime (high volatility regime)					
Model 2 AR(0)	1	0.996766829	0.9959647	0.985240586	0.99083889
Model 2 AR(1)		1	0.9946259	0.989951569	0.99668414
Model 3 AR(0)			1	0.99173637	0.99536928
Model 3 AR(1)				1	0.99480384
Model 4 AR(1)					1

Figure VI-24: The US stock market return series and Model 2-AR(1)'s smoothed posterior probability of observing a low volatility regime (3-regime)

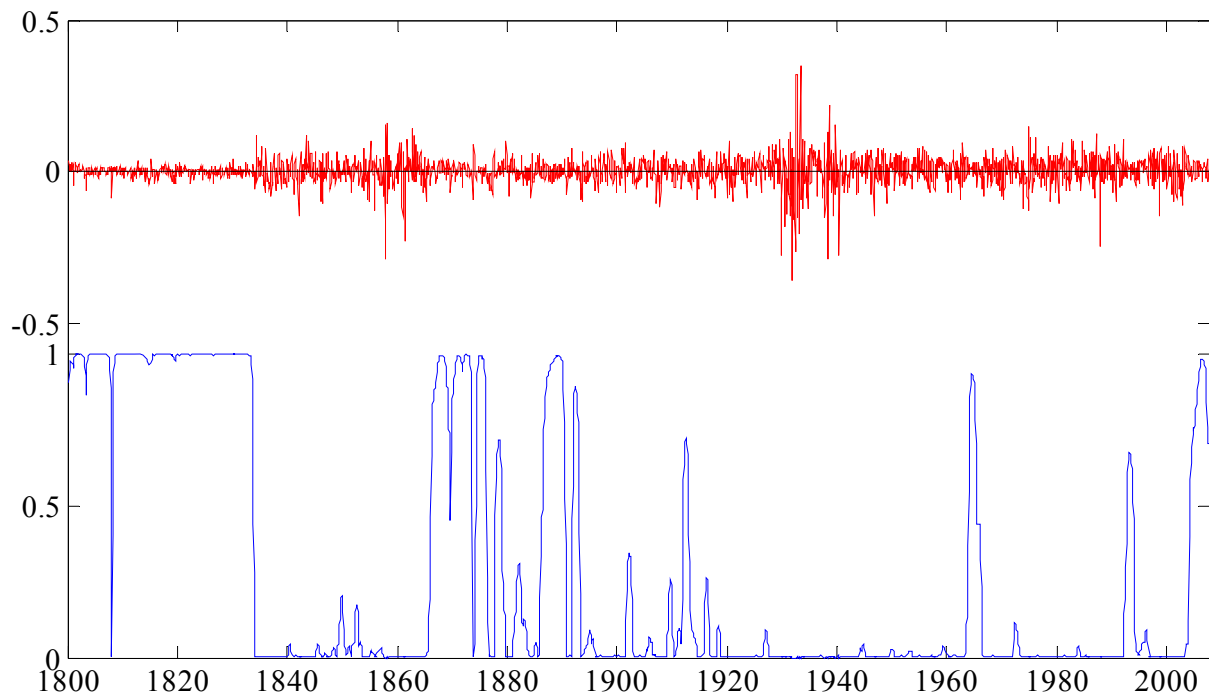
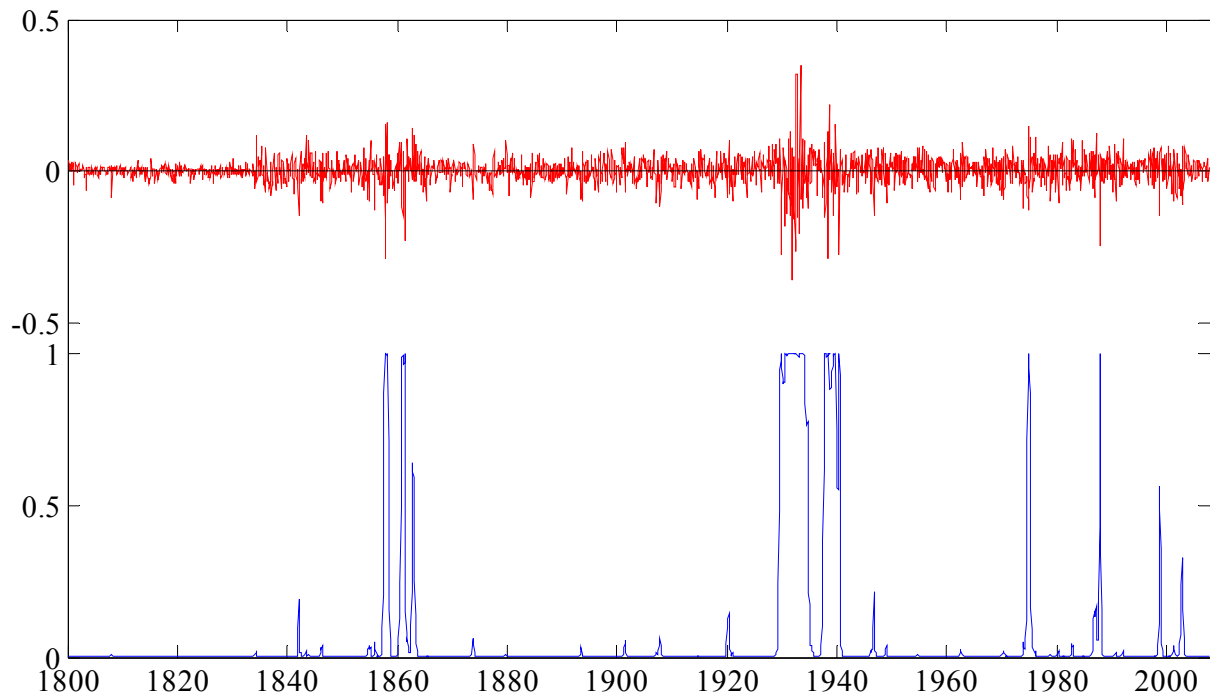


Figure VI-25: The US stock market return series and Model 2-AR(1)'s smoothed posterior probability of observing a high volatility regime (3-regime)



VI.5.7 Volatility feedback effect

The estimation results of the volatility feedback effect (Model 5) are presented in Table VI-23 (see Appendix VI). Notice that we've considered two different sets of information available to investors during each trading period. In Panel A of Table VI-23, we assume market investors observe past returns ($\Psi_{t-1} = \{r_{t-1}, r_{t-2}, \dots\}$) at the beginning of each trading period and make subjective guesses on the probabilities of experiencing a high volatility regime in the forthcoming trading month, based on the observed past returns. Panel B of Table VI-23 considers the situation that investors observe exactly which volatility regime the market was experiencing in the last trading month ($\Psi_{t-1} = \{S_{t-1}\}$). Since the parameter θ on $\Pr[S_{i2t} = 1 | \Psi_{t-1}]$ in equation (VI.5) represents the effect when market investors anticipate the return will draw from a high volatility regime given the available information to them before the current trading month, we would expect a positive θ in accordance with what the standard financial theory suggests - high volatility should be compensated by a higher expected return. To be specific, when a high volatility regime is highly expected (i.e. $\Pr[S_{i2t} = 1 | \Psi_{t-1}]$ is high), investors demand more expected return which results a significant and positive θ . On the other hand, if the current trading month's

high volatility regime was not anticipated at the beginning of the month (i.e. $\Pr[S_{i2t} = 1 | \Psi_{t-1}]$ is small), the average return between time $t-1$ and t will be dominated by μ_2 in equation (VI.5). Furthermore, if market investors are surprised by this unanticipated high volatility regime due to unfavourable information about fundamentals, a negative return (i.e. a negative μ_2) in the current period would be realized since the stock price must fall in order to obtain a higher expected return in the future to compensate for the high volatility in the future⁴². Similarly, if a low volatility regime appeared unexpectedly in the current trading month (i.e. $\Pr[S_{i2t} = 1 | \Psi_{t-1}]$ is high but $S_{i2t} = 0$), the average return between time $t-1$ and t will be higher than μ_1 .

The parameter estimates in Panel A, in general, are consistent with our prediction. The estimated θ parameters for all countries are significant and positive, which suggests that the volatility feedback effect is positive on anticipating a high volatility regime of the current trading month. Furthermore, since this θ parameter captures the effects of market volatility on all future discounted expected returns (and not just the contemporaneous expected return), it provides a more robust way in estimating a true relationship between market volatility and the expected return. On the other hand, the estimated negative μ_2 reveals the true relationship between market return volatility and realized returns. That is, having taken account for the volatility feedback effect, stock prices move in the opposite direction to the level of market volatility which creates an immediate realized negative stock return.

The magnitude of the estimated parameters θ and μ_2 in Panel B of Table VI-23, assuming market investors observe previous month's real volatility regime, is much deeper than those in Panel A. For example, the θ estimate for the US stock market assuming investors only observe past returns is 0.0196, while assuming investors are aware of the true market volatility regime in the previous month the magnitude of θ jumps to 0.0459. The negative relationship between market volatility and realized returns also deepens when using different information assumption, as shown in Panel B. The return of the US stock market, in the high volatility regime, changes from -0.0203 to -0.0477. This evident deepening of the estimated θ and μ_2 parameters, when we assume $\Psi_{t-1} = \{S_{t-1}\}$, suggests that the volatility

⁴² This relies on the assumption that investors are risk averse and volatility movements are persistent, i.e. a higher volatility today would lead to persistent upward movements in volatility in the future.

feedback effect signifies when investors are more certain about their guesses of the current market volatility regime. In other words, the better the information available to investors on predicting trading month's volatility regime, the more accurate the investors' prediction, and hence the higher the pressure on the stock prices to fall immediately.

VI.5.8 The co-movements (integration) of the stock markets

Next, we investigate the coherence in regimes over time for the six countries, which may shed some light on our understanding of the degree of co-movements (integration) of these countries' stock markets. Many recent studies have documented the trend of financial liberalization through time in many countries. The result of this increasing trend has seen many countries' capital markets are more integrated than before. For example, Hardouvelis, Malliaropulos and Priestley (2006) find that, in the second half of the 1990s, individual Euro-Zone country stock markets appear to be fully integrated into the EU market. Most studies in international capital market integration adopt methodologies similar to those used in Hardouvelis, *et al.* (2006), i.e. assume an International Asset Pricing Model proposed by Adler and Dumas (1983) and employ the Markov switching framework of Bekaert and Harvey (1995). Our analysis on the degree of integration of the six countries' stock market however, is based on the estimation results obtained in previous sections. Having established the fact that volatility feedback effect exists in all six countries' stock market (positive θ parameter estimates), we acquired the smoothed state probabilities of observing low and high volatility regimes as a by-product. Since the low and high volatility regimes identified in Model 5 are also associated with negative and positive stock market returns, respectively, the smoothed state probability at each point of time distinctively implies a full description of the stock market return's first two moments. By studying on the coherence in regimes of the six countries' stock market return, we could offer a different way to assess the co-movements of these markets over time.

Let's denote \Pr_{iHt} as the smoothed state probability of observing a high volatility regime in country i at time t . For each pair of countries, we consider the following 2×2 contingency table

	$\Pr_{1Ht} < 0.5$	$\Pr_{1Ht} \geq 0.5$
$\Pr_{2Ht} < 0.5$	H_{11}	H_{12}
$\Pr_{2Ht} \geq 0.5$	H_{21}	H_{22}

where the columns correspond to country 1's state probability of observing a high volatility regime either greater or equal to 0.5 or less than 0.5, while the rows correspond to country 2's state probability of observing a high volatility regime either greater or equal to 0.5 or less than 0.5. Hence, H_{11} and H_{22} correspond to co-movements in volatility regimes between country 1 and 2, while H_{12} and H_{21} correspond to the incoherence. At each time point, we will only observe one scenario among the four to occur. The co-movements between country 1 and 2 can be evaluated based on the ratio of the sum of the diagonal elements to the sum of all elements, which we call it coherence ratio. A coherence ratio tells us what proportion of time the two countries are simultaneously experiencing either a high volatility regime with low stock return or a low volatility regime with high stock return.

Table VI-22 reports the coherence ratio between pairs of countries over the full sample period. A higher value of the ratio means a higher coherence in regimes between the corresponding row country and column country. Interestingly, we see the highest coherence between two countries occurs on CAN-US relation with a coherence ratio of 0.8347, followed by AUS-UK with a ratio of 0.7568. Surprisingly, we find the coherence relation between SWI and US is also very high - the two countries' stock market returns have simultaneously experienced the same volatility regime in 73.34% of the time from March 1916 to September 2007. On the other hand, the weakest coherence relation occurs on CAN-UK relation, with a coherence ratio of 0.4596.

Table VI-22: Coherence in regimes between pairs of countries

	CAN	SWE	SWI	UK	US
AUS	0.6472	0.7049	0.6642	0.7568	0.6838
CAN		0.6602	0.7057	0.4596	0.8347
SWE			0.6642	0.6025	0.6984
SWI				0.5696	0.7334
UK					0.6632

In general, Table VI-22 offers us a good way to evaluate the co-movements between stock market returns in each pairs of countries. Yet, one may be interested to investigate the time evolution of the coherence in regimes between pairs of countries rather than the historical evaluation based on a static sample. In order to get the dynamic time evolution of the coherence ratio, we calculate the ratio starting from the beginning of the sample of each pairs of countries and sequentially enlarge the sample to the end of the sample period.

Therefore, at each time point, a new observation of coherence between two countries' stock market returns will increase the recalculated coherence ratio, while a detection of incoherence will decrease the ratio. In Figure VI-26, Figure VI-27 and Figure VI-28, we plot the time evolution of the coherence ratio between each pairs of countries. Generally, the coherence in regimes between Australian stock market and the rest of the countries' has decreased since the mid 90s. The most significant drops in the coherence in regime occurred between CAN-UK and SWI-UK relations, while an increase in coherence of regimes appeared in CAN-SWE and SWE-US relations.

Figure VI-26: Time evolution of the coherence in regimes between pairs of countries (1)

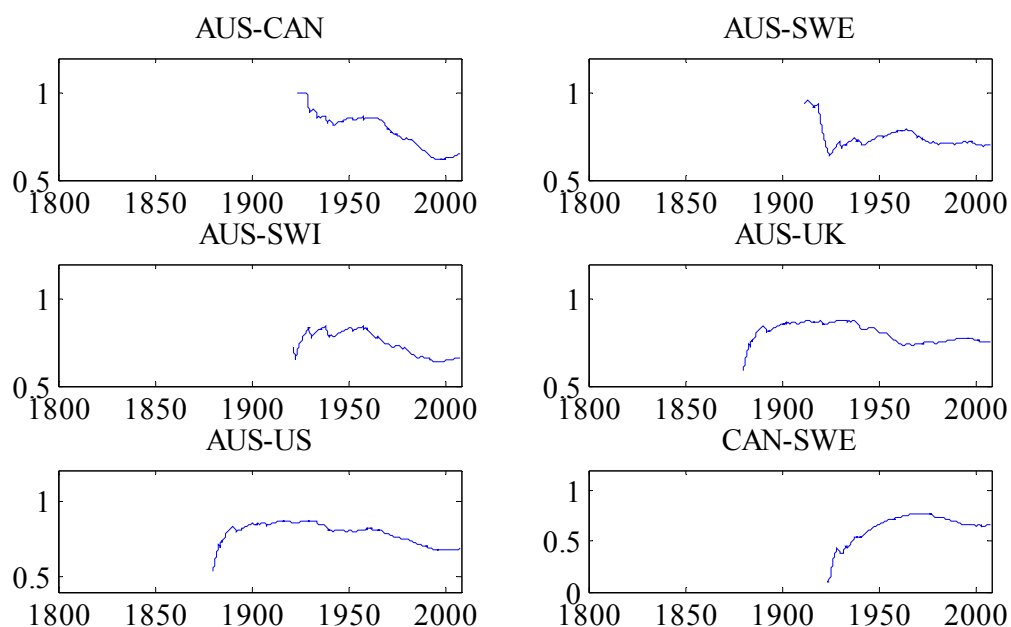


Figure VI-27: Time evolution of the coherence in regimes between pairs of countries (2)

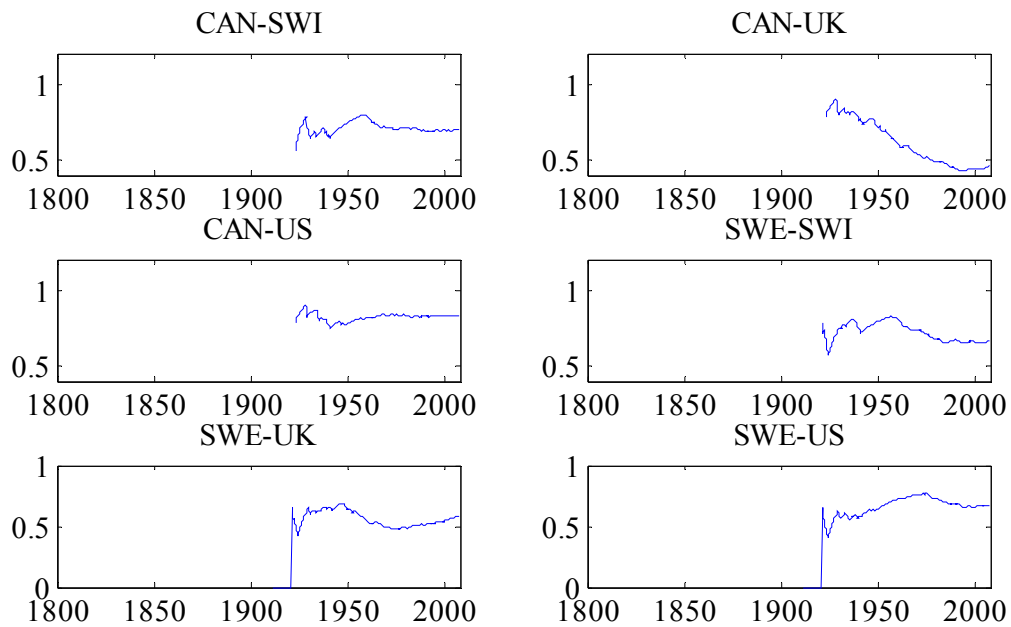
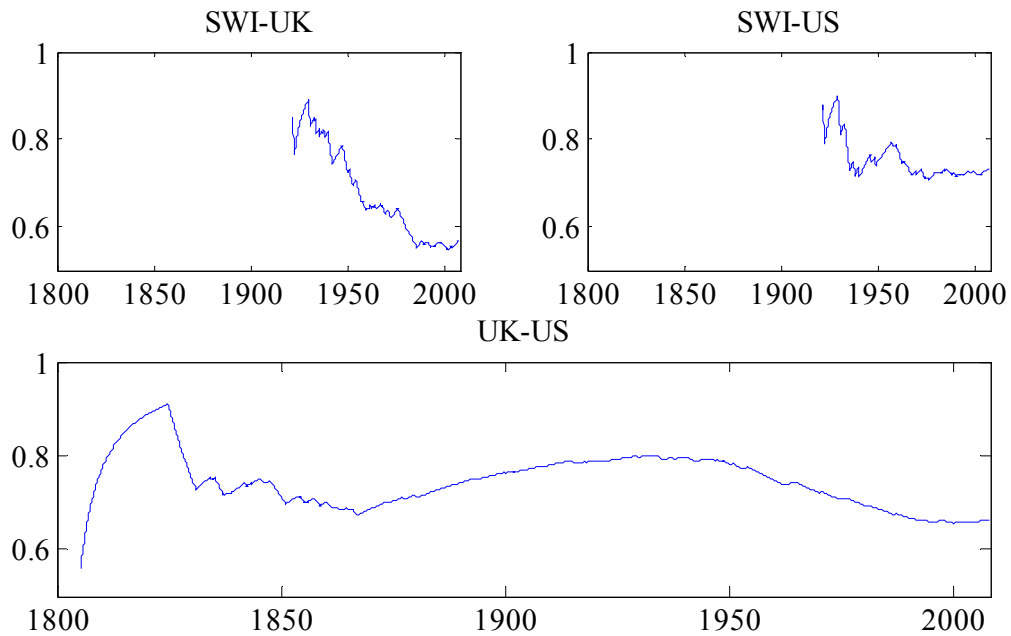


Figure VI-28: Time evolution of the coherence in regimes between pairs of countries (3)



VI.6 Conclusion

In this study, we investigate the time-varying risk return relation for six countries using aggregate stock market data span century-long time. The models we used are flexible Markov

switching models with different assumptions on the regime dependence of mean and volatility of the stock returns. We find that the Markov switching models assuming both regime dependent mean and volatility with a 3-regime specification are more capable to capture the extreme movements of the stock market which are short-lived, which helps us to capture the stock return's leptokurtosis in its distribution. In addition, as most of the extreme movements in the stock market are downwards, Markov switching models with a 3- regime specification would better capture the negative skewness in the stock returns than their 2-regime counterparts.

The volatility feedback effect that we studied on these six countries shows a positive sign on anticipating a high volatility regime of the current trading month. Since the volatility feedback effect captures the effects of market volatility on all future discounted expected returns (and not just the contemporaneous expected return), it provides a more robust way in estimating a true relationship between market volatility and the expected return. On the other hand, the estimated negative μ_2 reveals the true relationship between market return volatility and realized returns. That is, having taken account for the volatility feedback effect, stock prices move in the opposite direction to the level of market volatility which creates an immediate realized negative stock return.

The investigation on the coherence in regimes over time for the six countries shows different results for different pairs of countries. While some countries show increasing coherence in regimes, like the SWE-CAN and SWE-US pairs, many countries show decreasing co-movements in volatility regimes (e.g. the SWI-UK and SWI-US). Given the evidence of decoupling in coherence of volatility regime over time, future research on the implication of this effect on global portfolio diversification and investment strategy is promising.

Bibliography

Adler, Michael, and Bernard Dumas, 1983, International portfolio choice and corporation finance: A synthesis, *Journal of Finance* 38, 925-984.

- Bae, J., C. J. Kim, and C. R. Nelson, 2007, Why are stock returns and volatility negatively correlated?, *Journal of Empirical Finance* 14, 41-58.
- Baillie, Richard T., and Ramon P. DeGennaro, 1990, Stock returns and volatility, *Journal of Financial and Quantitative Analysis* 25, 203-214.
- Bansal, R., and C. Lundblad, 2002, Market efficiency, asset returns, and the size of the risk premium in global equity markets, *Journal of Econometrics* 109, 195-237.
- Bekaert, Geert, and Campbell R. Harvey, 1995, Time-varying world market integration, *Journal of Finance* 50, 403-444.
- Bekaert, Geert, and Guojun Wu, 2000, Asymmetric volatility and risk in equity markets, *Review of Financial Studies* 13, 1-42.
- Black, F., 1976, Studies of stock price volatility change, *In: Proceedings of the American Statistical Association, Business and Economic Statistics Section* 177-181.
- Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307-327.
- Bollerslev, T., R. Y. Chou, and K. F. Kroner, 1992, Arch modeling in finance - a review of the theory and empirical evidence, *Journal of Econometrics* 52, 5-59.
- Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge, 1988, A capital asset pricing model with time-varying covariances, *The Journal of Political Economy* 96, 116-131.
- Bollerslev, Tim, Julia Litvinova, and George Tauchen, 2006, Leverage and volatility feedback effects in high-frequency data, *Journal of Financial Econometrics* 4, 353-384.
- Brandt, M. W., and Q. Kang, 2004, On the relationship between the conditional mean and volatility of stock returns: A latent var approach, *Journal of Financial Economics* 72, 217-257.
- Breen, William, Lawrence R. Glosten, and Ravi Jagannathan, 1989, Economic significance of predictable variations in stock index returns, *Journal of Finance* 44, 1177-1189.
- Campbell, J. Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373-399.
- Campbell, J. Y., and L. Hentschel, 1992, No news is good news - an asymmetric model of changing volatility in stock returns, *Journal of Financial Economics* 31, 281-318.
- Campbell, J. Y., and R. J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195-228.
- Chou, Ray Yeutien, 1988, Volatility persistence and stock valuations: Some empirical evidence using garch, *Journal of Applied Econometrics* 3, 279-294.
- Christie, A. A., 1982, The stochastic behavior of common stock variances - value, leverage and interest rate effects, *Journal of Financial Economics* 10, 407-432.
- Engle, Robert F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica: Journal of the Econometric Society* 50, 987-1007.
- Engle, Robert F., David M. Lilien, and Russell P. Robins, 1987, Estimating time varying risk premia in the term structure: The arch-m model, *Econometrica: Journal of the Econometric Society* 55, 391-407.
- Engle, Robert F., and Victor K. Ng, 1993, Measuring and testing the impact of news on volatility, *Journal of Finance* 48, 1749-1778.
- French, K. R., G. W. Schwert, and R. F. Stambaugh, 1987, Expected stock returns and volatility, *Journal of Financial Economics* 19, 3-29.
- Garcia, R., 1998, Asymptotic null distribution of the likelihood ratio test in markov switching models, *International Economic Review* 39, 763-788.
- Ghysels, E., P. Santa-Clara, and R. Valkanov, 2005, There is a risk-return trade-off after all, *Journal of Financial Economics* 76, 509-548.

- Glosten, Lawrence R. , Ravi Jagannathan, and David E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779-1801.
- Hamilton, J. D., and R. Susmel, 1994, Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics* 64, 307-333.
- Hansen, B. E., 1992, The likelihood ratio test under nonstandard conditions: Testing the markov switching model of gnp, *Journal of Applied Econometrics* 7, S61-S82.
- Hardouvelis, Gikas A., Dimitrios Malliaropulos, and Richard Priestley, 2006, Emu and european stock market integration, *Journal of Business* 79, 365-392.
- Harrison, Paul, and Harold H. Zhang, 1999, An investigation of the risk and return relation at long horizons, *Review of Economics and Statistics* 81, 399-408.
- Harvey, C. R., 1989b, Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics* 24, 289-317.
- Jegadeesh, Narasimhan, 1991, Seasonality in stock price mean reversion: Evidence from the u.S. And the u.K, *Journal of Finance* 46, 1427-1444.
- Kim, Chang-Jin, James C. Morley, and Charles R. Nelson, 2004, Is there a positive relationship between stock market volatility and the equity premium?, *Journal of Money, Credit & Banking (Ohio State University Press)* 36, 339-360.
- Lintner, John, 1965, Security prices, risk, and maximal gains from diversification, *Journal of Finance* 20, 587-615.
- Ludvigson, S. C., and S. Ng, 2007, The empirical risk-return relation: A factor analysis approach, *Journal of Financial Economics* 83, 171-222.
- Lundblad, C., 2007, The risk return tradeoff in the long run: 1836-2003, *Journal of Financial Economics* 85, 123-150.
- Mayfield, E. S., 2004, Estimating the market risk premium, *Journal of Financial Economics* 73, 465-496.
- Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica: Journal of the Econometric Society* 41, 867-887.
- Nelson, Daniel B., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica: Journal of the Econometric Society* 59, 347-370.
- Schwert, G. William, 1989, Why does stock market volatility change over time?, *Journal of Finance* 44, 1115-1153.
- Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425-442.
- Turner, C. M., R. Startz, and C. R. Nelson, 1989, A markov model of heteroskedasticity, risk, and learning in the stock market, *Journal of Financial Economics* 25, 3-22.
- Whitelaw, Robert F., 1994, Time variations and covariations in the expectation and volatility of stock market returns, *Journal of Finance* 49, 515-541.

Appendix VI

Table VI-23: Estimation results of Model 5 for each country

	Panel A: $\Psi_{t-1} = \{r_{t-1}, r_{t-2}, \dots\}$						Panel B: $\Psi_{t-1} = \{S_{t-1}\}$					
	AUS	CAN	SWE	SWI	UK	US	AUS	CAN	SWE	SWI	UK	US
μ_1	0.00506 (0.00091)	0.00691 (0.00147)	0.00665 (0.00163)	0.00464 (0.00160)	0.00188 (0.00065)	0.00273 (0.00099)	0.00560 (0.00087)	0.00687 (0.00145)	0.00482 (0.00154)	0.00491 (0.00169)	0.00286 (0.00061)	0.00223 (0.00106)
μ_2	-0.0047 (0.00408)	-0.0324 (0.01132)	-0.0025 (0.00687)	-0.0155 (0.00695)	-0.0009 (0.00331)	-0.0203 (0.00592)	-0.0067 (0.00737)	-0.0680 (0.02145)	-0.0647 (0.01242)	-0.0314 (0.01642)	-0.02478 (0.00633)	-0.0477 (0.00913)
σ_1	0.01982 (0.00070)	0.03206 (0.00122)	0.02997 (0.00117)	0.02725 (0.00108)	0.01614 (0.00047)	0.02823 (0.00080)	0.01968 (0.00069)	0.03175 (0.00104)	0.03082 (0.00104)	0.02748 (0.00116)	0.01626 (0.00048)	0.02741 (0.00110)
σ_2	0.06165 (0.00253)	0.07830 (0.00543)	0.06969 (0.00359)	0.06559 (0.00356)	0.05668 (0.00155)	0.08460 (0.00479)	0.06103 (0.00244)	0.07626 (0.00499)	0.07435 (0.00440)	0.06624 (0.00390)	0.05691 (0.00155)	0.07990 (0.00448)
p_{11}	0.96332 (0.00897)	0.97634 (0.00851)	0.97996 (0.00661)	0.95886 (0.01243)	0.97905 (0.00485)	0.97429 (0.00582)	0.96483 (0.00877)	0.97825 (0.00787)	0.97942 (0.00602)	0.95839 (0.01260)	0.98169 (0.00441)	0.96801 (0.00778)
p_{22}	0.91587 (0.02175)	0.87975 (0.03606)	0.95763 (0.01597)	0.90326 (0.03419)	0.96526 (0.00807)	0.88156 (0.02265)	0.92386 (0.02012)	0.89288 (0.03655)	0.93814 (0.02263)	0.89332 (0.03877)	0.96789 (0.00751)	0.85189 (0.02978)
α_1	0.11283 (0.03217)	0.09442 (0.03483)	0.14151 (0.04123)	0.11569 (0.03677)	0.22820 (0.02816)	0.15313 (0.02392)	0.11439 (0.03258)	0.11390 (0.03704)	0.15780 (0.03613)	0.11206 (0.03655)	0.22290 (0.02648)	0.15855 (0.02507)
α_2	0.03676 (0.04780)	0.23495 (0.07904)	0.18813 (0.05016)	0.15956 (0.05928)	0.16108 (0.03267)	0.09170 (0.05058)	0.03835 (0.04684)	0.22008 (0.07616)	0.18121 (0.05650)	0.16015 (0.06164)	0.16408 (0.03283)	0.08687 (0.04975)
θ	0.00765 (0.00382)	0.02093 (0.00855)	0.00295 (0.00572)	0.01391 (0.00647)	0.00292 (0.00257)	0.01960 (0.00529)	0.00845 (0.00740)	0.04922 (0.01827)	0.05715 (0.01045)	0.02608 (0.01529)	0.0200 (0.00532)	0.04591 (0.00874)
Log L	3282.561	1941.814	2147.777	2020.148	5409.549	4717.542	3281.862	1941.489	2151.086	2019.298	5410.889	4716.433
Q(12)	19.095 [0.086]	14.219 [0.287]	40.440 [0.000]	30.692 [0.002]	24.529 [0.017]	21.106 [0.049]	19.116 [0.086]	13.126 [0.360]	42.081 [0.000]	29.984 [0.003]	25.730 [0.012]	21.836 [0.039]
Q ² (12)	32.260 [0.001]	10.665 [0.558]	30.329 [0.002]	8.167 [0.772]	200.474 [0.000]	89.009 [0.000]	16.015 [0.191]	9.015 [0.702]	34.578 [0.001]	8.026 [0.783]	185.455 [0.000]	119.372 [0.000]

Note: This Ljung-Box Q statistics is a modified portmanteau test of the original one, which allows us to test the ARCH effect of the residuals. Q(12) refers to the Ljung-Box test for no serial correlation up to 12 lags. Q²(12) refers to the Ljung-Box test for no ARCH effect up to 12 lags. The p-values are shown in square brackets. A smaller p-value indicates a rejection of the null hypothesis of no serial correlation or no ARCH effect.

Matlab Code (Model 5)

The following codes require CompEcon Toolbox to run. The toolbox can be downloaded from www4.ncsu.edu/~pfackler/compecon

```
%%=====us_m5_2s_run.m=====
% us model 5(volatility feedback effect, switching mean and variance, and
% switching phi) with 2 states
%=====

%% START
clear all;
close all;

%% GLOBAL VARIABLES
global yy xx tt slt_mat s2t_mat

%% LOAD DATA AND PREPARE RETURN SERIES
cd ..
data=load('US_Equity.txt');
us_index=data;
us_index_t=size(us_index,1);
% create monthly return series and lagged monthly return series
us_rt=log(us_index(2:end))-log(us_index(1:end-1));
us_rt_t=size(us_rt,1);
% return data now starts from Feb1800 to Sep2007
us_rt_lag=us_rt(1:end-1);
us_rt_lead=us_rt(2:end);
cd('model_5');

% define variables used in estimation
yy=us_rt_lead;
xx=us_rt_lag;
tt=size(yy,1);
% the date that will be used in plots(Mar1800 to Sep2007)
date_tt=3/12+1800:1/12:9/12+2007;

%% SETTING INITIAL VALUES FOR PARAMETERS

%para_in=[0.2;0.5;0.34;0.1;-0.78;-0.123;0.1;0.5;0.214;0.5497;0.4578;0.1265;0.7984;0.2;0.3];
para_in=[-5.42440225755167;-2.82925367446647;
0.686212861936474;0.3513405962944;
-0.239493770655378;0.0268833329977971;0.58403217546369;0.5;-0.005];
%% CALL MS FILE TO GET ST_MAT, S1T_MAT, S2T_MAT AND S3T_MAT
ar_lag=1;
num_of_states=2;
[st_mat,s1t_mat,s2t_mat]=us_m5_2s_ms(ar_lag, num_of_states);

%% ESTIMATION
options=optimset('Display','iter','TolX',1e-4,'TolFun',1e-8,...
'MaxFunEvals',100000,'MaxIter',100000);
[xout,fval,exitflag,output]=fminsearch(@us_m5_2s_likhfcn,para_in,options);
xfnl=us_m5_2s_trans(xout);
% S.E. and t-test
h_0=fdhess(@us_m5_2s_likhfcn,xout); % calculate Hessian matrix

g_0=fdjac1(@us_m5_2s_trans,xout,[]); % calculate gradient
```



```

h_fnl=g_0*(inv(h_0))*g_0'; % final Hessian
std_fnl=sqrt(diag(h_fnl)); % S.E.
t_ratio=xfnl/std_fnl;

%%% filter
[var_mat,fit_mat,prtt,prtl,sd_resid]=us_m5_2s_filter(xout);

%%% smooth
prob_smooth_1=us_m5_2s_smooth(prtt,prtl,xout);

function loglik_value=us_m5_2s_likhfcn(para_in)
%%% GLOBAL VARIABLES
global yy xx tt s1t_mat s2t_mat

%%% ASSIGN PARAMETERS
para=us_m5_2s_trans(para_in);

% variance
sig1=para(1);
sig2=sig1+para(2);

% transition probs
prob_tran=[para(3)    1-para(4);
           1-para(3)  para(4) ];

% mu and phi
mu1 =para(5);
mu2 =para(6);

phi1 =para(7);
phi2 =para(8);

theta=para(9);
%%% THE VARIANCE AT TIME T AND MU AT TIME (T-1) AND (T)
var_t=sig1^2*s1t_mat(:,2)+sig2^2*s2t_mat(:,2); % 4 by 1
mu_mat=mu1*s1t_mat+mu2*s2t_mat; % 4 by 2, first col is S(t-1)
phi_mat=phi1*s1t_mat(:,2)+phi2*s2t_mat(:,2);

%%% STEADY STATE PROBS AS INITIAL PROBS

A=[eye(2)-prob_tran; ones(1,2)];
B=[0;0;1];
ss_prob=inv(A'*A)*A'*B; % steady state probabilities, 2 by 1

%%% START ITERATION
loglik_value=0;

for j_iter=1:tt

    vec_ss_prob=reshape([ss_prob';ss_prob'],4,1); % 4 by 1
    vec_prob_tran=reshape(prob_tran,4,1); % 4 by 1
    vec_prob_joint=vec_ss_prob.*vec_prob_tran; % 4 by 1, Pr[St(-1), St(0)|I(0)]
    prob_TL1=vec_prob_joint(2)+vec_prob_joint(4);

    er=(yy(j_iter)*ones(4,1)-phi_mat*xx(j_iter))-(mu_mat(:,2)+theta*prob_TL1-phi_mat.*mu_mat(:,1));

```

```

lik=(1./sqrt(2*pi.*var_t)).*exp(-0.5*(er.*er)./var_t)...
.*vec_prob_joint;
lik_value=sum(lik);

update_prob=lik/lik_value;

loglik_value=loglik_value-log(lik_value);

ss_prob=update_prob(1:2)+update_prob(3:4); % 2 by 1
end

end% of function

% us_m4_2s_ms.m
function [st_mat,s1t_mat,s2t_mat]=us_m5_2s_ms(ar_lag, num_of_states)
% this file generates markov switching variables for us model 5 with 2
% states
%-----
num_of_st=ar_lag+1; % one st_lag1 and st_current
%num_of_states=3;
dmnsion=num_of_states^num_of_st;
st_mat=zeros(dmnsion,num_of_st);

i=1;
for st_lag1=1:num_of_states
    for st=1:num_of_states
        st_mat(i,:)=st_lag1 st;
        i=i+1;
    end
end

s1t_mat=zeros(dmnsion,num_of_st);
s2t_mat=zeros(dmnsion,num_of_st);

for i=1:dmnsion
    for j=1:num_of_st
        if st_mat(i,j)==1
            s1t_mat(i,j)=1;
        end
    end
end

for i=1:dmnsion
    for j=1:num_of_st
        if st_mat(i,j)==2
            s2t_mat(i,j)=1;
        end
    end
end

function temp=us_m5_2s_smooth(pr_tt,pr_tl,para_in)
global tt

```

```

para=us_m5_2s_trans(para_in);

pr_tr=[para(3)    1-para(4);
        1-para(3)    para(4) ];

p11=pr_tr(1,1); p12=pr_tr(2,1);
p21=pr_tr(1,2); p22=pr_tr(2,2);


pr_tt1=pr_tt(:,1);
pr_tt2=pr_tt(:,2);


pr_tl1=pr_tl(:,1);
pr_tl2=pr_tl(:,2);


pr_sm1=pr_tt1;
pr_sm2=pr_tt2;


for j_iter=tt-1:-1:1
    pr_sm11=pr_sm1(j_iter+1,1)*p11*pr_tt1(j_iter,1)/pr_tl1(j_iter+1,1);
    pr_sm12=pr_sm2(j_iter+1,1)*p12*pr_tt1(j_iter,1)/pr_tl2(j_iter+1,1);

    pr_sm21=pr_sm1(j_iter+1,1)*p21*pr_tt2(j_iter,1)/pr_tl1(j_iter+1,1);
    pr_sm22=pr_sm2(j_iter+1,1)*p22*pr_tt2(j_iter,1)/pr_tl2(j_iter+1,1);

    pr_sm1(j_iter,1)=pr_sm11+pr_sm12;
    pr_sm2(j_iter,1)=pr_sm21+pr_sm22;

end

temp=[pr_sm1 pr_sm2];
end

```

Chapter VII A Markov Switching Unobserved Component Analysis of the CDX Index Term Premium

VII.1 Introduction

The sub-prime mortgage crisis, unveiled in July 2007, has caused remarkable losses in the credit markets. A large number of systemically important financial institutions had been forced to write off mortgages and related securities linked to credit derivatives instruments, like credit default swaps (CDSs) and collateralized debt obligations (CDOs). Great uncertainties filled almost every corner of the financial markets, which seriously interrupted its normal functioning (see Taylor and Williams (2009)). It has been argued (see, amongst the others, Stulz (2009)) that the 2007/2008 sub-prime crisis was amplified through structured credit products-tranches trading. Consequently, if on the one hand, these instruments seem to have enriched the scope of investment strategies, on the other hand, their increased complexity depth have unduly induced instability in financial markets.

CDS indices are simply portfolios of single name default swaps, serving both as trading vehicles and as barometers of credit market conditions. Users of the most popular indices (the Dow Jones CDX North American investment grade (CDX-IG) and the iTraxx Europe investment grade (iTraxx-IG)) include those who want to hedge against credit defaults of pooled entities and those who want to speculate. These indices are responsible for the increased liquidity and popularity of tranching of credit risk. By buying protection on an index, an investor is protected against defaults in the underlying portfolio and makes quarterly premia payments to the protection seller. If there is a default, the protection seller pays par to the protection buyer.

The term premium of the CDX index, which is measured as the difference between a longer maturity CDX index series and a shorter maturity one, e.g. the difference between the CDX 10-Year index and the CDX 5-Year index, can be viewed as representing the 5-year forward uncertainty regarding corporate default after the next 5 years. Accordingly, the CDX term premium can be interpreted as an early warning market indicator of improvement or deterioration in macroeconomic conditions 5 years hence. If an investor perceives the difference between the 5-year index spread and the 10-year spread too steep, in other words,

that the implied probability of default⁴³ between 5 and 10 years is higher than that implied from fundamentals, but he/she expects the slope to flatten, then this investor could buy 5-year protection and sell 10-year protection on the CDX index. Finance theory suggests that the credit curves of companies with high credit quality should be upward sloping, whereas those of companies having very poor credit quality do exhibit negative slopes. For example, the credit risk of an AAA rated corporate bond should in general be positively correlated with its maturity, and hence the required yields slope upwards against its maturity. In cross-sectional space, the likelihood of credit quality deterioration should increase as rating lowers, which is to say that the required average yield should increase with the downgrading of corporate bonds. However, the credit curve for a company on the brink of default (or with foreseeable immediate downgrade) would invert to trend negatively to reflect higher credit risk in the near future. As a result, the yield for such bond is very high for short maturities but relatively lower for longer maturities, which reflects investors' view that it is still possible for this company to improve its credit quality for longer term maturities.

CDX curve trading has assumed enormous importance over the latest very turbulent period. Clearly, curves tend to flatten in periods of imminent higher default rates, and tend to be steep in periods of economic expansion.

Index curve trading is generally motivated on one or more of the following⁴⁴

- a) As a way of expressing market direction views with different risk-reward profiles.
- b) Carry and roll-down reasons.
- c) Hedging purposes – both cash and CDS underlying portfolios.

Many opportunities for trading curves on single-name CDS occur around forecasted or announced specific corporate actions. Such events change the perception of a company's creditworthiness and the shape of the CDS curve also evolves. Curve trades can be more attractive than outright positions around events, thanks to the variation in the available payoff profiles.

⁴³ The implied probability of default here is a risk-neutral concept, i.e. it is jointly determined by actual default probability and the risk premium, both default risk premium and liquidity risk premium. As a result, the level and changes of the CDX term premium can be driven by the forward uncertainty regarding corporate defaults (a physical concept) as well as the forward uncertainty regarding risk premium.

⁴⁴ For more details, see Barclays Capital Research (2008b)

In addition, macroeconomic conditions can trigger default events that affect the curves of not only specific entities but also of entire industries. Changes in consumer preferences, the monetary policy stance, and developments in the housing market are critical industry-wide events and market sentiment often transcend worries about profitability and focus instead on viability and the possibility of default of a specific firm. In this case, recovery expectations, following a higher default rate regime, become progressively important in determining the curve shape of the index, as it clearly tends to flatten.

In this paper, we investigate the dynamic behavior of the CDX index term premium through time by using a Markov Switching Unobserved Component (MS-UC) model. In the econometric literature, several approaches have been proposed on how the univariate time series could be decomposed. A well-established methodology is the unobserved components approach, postulated in separate contributions by Harvey (1985), Watson (1986) and Clark (1987), respectively. It seems natural to consider an economic time series in terms of permanent and stationary components. The decomposition of a univariate time series into these two components is a primary tool for analyzing business cycles, with these two components often used as measurements of unobserved trend and cycle. Researchers also use unobserved component models to study the mean reversion in stock prices. Fama and French (1988) find a stationary mean reverting component in addition to a permanent component in the US stock price dynamics. Poterba and Summers (1988) test the existence of a stationary component although does not perform a formal decomposition of the stock prices in stationary and permanent components. Others like Lo and MacKinlay (1988) and Kim, Nelson and Startz (1991) use variance ratio tests to detect mean reversion in stock prices. Although the evidence of mean reversion in stock prices is mixed, as Summers (1986) argues, statistical tests used in testing the random walk hypothesis have usually low power against the alternative of mean reversion.

In formulating an unobserved components model for econometric analysis, we depart from others working on the observable determinants of CDS indices. Alexander and Kaeck (2008) and Byström (2006), for example, relate the CDS/CDX spread to several observed variables (such as the slope of the yield curve, stock market return and stock market volatility), and analyze the significance of each observable variable in determining the CDS iTraxx Europe spread using single-equation regression. Our interest in this paper, however, is to study how the factors themselves (not the factor loadings) drive the dynamics of the term

premium. Since the CDX index measures the economy-wide default probabilities (the higher the index value, the higher the probability of default on firms included in the index), the macroeconomic conditions, which can be encompassed by those fundamental factors, will be closely related to the CDX index value and its term premium. To characterize the observed patterns of volatility jumps on the CDX index term premium, we allow on the innovation terms a regime switching process, following two distinct first-order Markov chain variables.

This paper contributes to the rapidly growing literature on structured credit in its attempt to understand the evolution of the term premium of the CDS indices market and its link to observed macroeconomic and financial information. Our paper has two main contributions. First, we present a readily implementable new approach to modeling CDS index dynamics, by conducting a regime dependent factor analysis of the evolution of the CDX index. Second, we provide insights into how the fundamental and volatility components of the CDX index are determined by daily observed monetary policy and stock market variables, before and after the onset of the global financial crisis.

The current literature on CDS is primarily limited to the pricing with a large strand of it revolving around the key determinants of these contracts. First, there is an extensive literature on the driving forces of CDS spread ranging from the model of Hull, Predescu and White (2004), Aunon-Nerin, Cossin, Hricko and Huang (2002), which examine the relationship between CDS spread and credit spreads, to more elaborate analysis by – amongst the others – Zhu (2006), Longstaff, Mithal and Neis (2005) and Blanco, Simon and Marsh (2005) which include also bond and equity markets measures.

Much of the research on credit markets has focused on corporate bond spreads and single-name CDS spread. Despite a sizeable literature on credit risk empirical studies on CDS that involve the modeling of the entire credit curve are uncommon. A major reason for this is that data on the CDS spread for a wide range of maturities have only recently become available. Consequently there is a paucity of empirical works regarding CDS indices, with studies focused mainly on the North America CDX investment grade index (CDX-IG).

Our work is also closely related to two recent studies by Pan and Singleton (2008) and Zhang (2008), who attempt to estimate default risk using the entire credit curve of sovereign CDS spread. Byström (2005), Byström (2006) and Alexander and Kaeck (2008) are the early studies on CDS indices. In a correlation study of a sample of European CDS

iTraxx indices for different industrial sectors, Byström (2005) finds a tendency for iTraxx spread to narrow when stock prices rise, and vice versa. Furthermore, he finds that stock market reacts quicker than the iTraxx market to firm-specific information and the stock price volatility is significantly and positively related to the volatility of CDS spread. Alexander and Kaeck (2008) use a Markov switching model to examine the determinants of the European iTraxx index in two different regimes. They find the CDS market is sensitive to stock returns under ‘ordinary’ market conditions but extremely sensitive to stock volatility during turbulent periods. One recent paper by Bhar, Colwell and Wang (2008), which is mostly related to our paper, decomposes three European CDS iTraxx indices spread into persistent and stationary components using the Kalman filter. They investigate these dynamics for two different maturities (5 and 10 years) and find that the stationary component is affected largely by stock market volatility whereas the persistent component is more sensitive to illiquidity. However, their sample period does not include the recent sub-prime mortgage crisis. Therefore, the dynamic behavior of these two components during crisis time remains still unexplained.

Blanco, *et al.* (2005) analyze the relationship between investment grade bonds and CDS, and explore the determinants of CDS spread. They find that the theoretical relationship linking credit spreads and CDS spread holds reasonably well for most of the investment grade reference entities. In addition, they report that increases in interest rates and equity prices reduce CDS spread whilst a steeper-sloping yield curve has the opposite effect.

The paper’s main results are as follows. First, the inversion of the CDX term premium is induced by sudden changes in the stationary component, which represents the evolution of the fundamentals underpinning the probability of default in the economy. Equally notable is that our findings show that the non-stationary component, which represents increases in volatility, spikes quite dramatically around the occurrence of tail risk events (e.g. Bear Sterns bailout and Lehman Brothers bankruptcy).

Second, the empirical evidence strongly suggests that the direct impacts of monetary policy rates and the slope of the yield curve on the term premium of CDX index are time varying and depending on business cycle. Credit risk modeling that ignores this regime dependent feature would bias the pricing of credit contracts. Developments in both the first and second moments of the equity market have a lasting influence on both components, with more pronounced effects in volatile market conditions.

The paper is organized as follows. Section 2 discusses the possible economic determinants of the term premium and suggests its decomposition into two unobserved components allowing for regime switching. Section 3 presents and discusses the data used in the estimation. The results are reported in Section 4 and Section 5 concludes.

VII.2 Motivation and Methodology

The econometric methodology employed in this paper is based on the statistical approach developed initially by Nervole, Grether and Carvalho (1979) and developed further by Harvey (1989a) and Harvey and Shephard (1993). The essential element of this methodology is to estimate a model which considers the observed time series as being the sum of permanent and stationary components. These components capture the salient features of the series that may be unobserved and are useful in explaining and predicting its time evolution. The Kalman filter is employed as the most efficient means of updating the state as new information becomes available, in linear models.

Although the latent variable model is an effective tool in decomposing macro/financial variables into a number of unobservable components, the usefulness of the model is however still limited if we are unable to link the components to a set of observable economic variables. To overcome this problem, one may model the unobserved components and observed variables together in a macro-finance setting (as suggested, for example, by Ang and Piazzesi (2003)). Our analytical approach here is instead as follows: we begin by filtering out the unobserved components and then in a second step, we empirically estimate the relationship between the unobserved components and a set of variables observed at the same frequency.

Our aim is to test for the economically meaningful relationship between the unobserved components and a set of observed information that is available to both market participants and policy makers. Such link, if established, will add predictive ability to the model as the evolution of the components will be conditional on data and will enhance the model's analytical appeal.

At a conceptual level, the US Federal Fund Rate is the standard monetary policy tool available to the Fed to influence the short end of the yield curve and hence, in turn, affects investors' expectations on the movements of long-term interest rates. An increase in Federal Fund Rate signals the Fed's reaction against the risk of rising inflation in the near future and

will aggravate the external financing position of companies that rely heavily on short-term financing. The impact of monetary policy on the term premium will be conditional on the state of the economy. Under “normal” conditions a tightening of monetary policy will indicate future inflationary pressures due to expanding demand. In this case, increases in the policy rate may be consistent with mitigating insolvency risks and thus with lower 5-year CDX premium. However in periods of crisis when expectations of future demand are gloomy, the same rate increases will enhance the probability of imminent default as the companies’ abilities to secure funds at reasonable rates are reduced, resulting in widening 5-year CDX premium and a flattening of the CDX index term premium.

The slope of the yield curve reflects simply a forward expectation of how the short-term interest rate is expected to fluctuate over a long-term horizon and is largely driven by the market-wide expectations about the future path of monetary policy. Once more the impact of changes in the yield curve will depend upon the prevailing market conditions. Under stable market conditions increases in the long-rate imply future rises of the short rate. Such predicted evolution will impact positively on both the 5 and the 10-year CDX spread, rendering ambiguous its effect on the term premium. Under crisis condition the ‘steepening’ due to decreases in the short-rate curve will reduce the 5-year spread relatively to the 10-year, increasing the term premium of the CDX index, as the reduction in the short-rate reduces the probability of imminent default.

Companies’ borrowing depends largely on the market value of their net worth (financial and tangible assets). Asymmetric information between borrowers and lenders, would prompt lenders to set forth the abilities of borrowers to repay the debt, which will take the form of collateralizing their financial assets. Falling asset prices erode the value of collateral, tightening credit and depressing demand. Through the so-called “credit channel”, the level of economic activity and the aggregate output will eventually shrink. If an adverse shock to the macro-economy is amplified by credit rationing, conditions in the real economy and in financial markets mutually reinforce each other, giving rise to a feedback loop which may lead to a deep recession. This self-reinforcing process, known as the “*financial accelerator*” (a term coined by Bernanke (1981), Bernanke (1983), Bernanke and Gertler (1989), and Bernanke, Gertler and Gilchrist (1996)), operates in reverse during a downturn. Increases of the equity index return will always result in reduction in the 5-year CDX spread as the firm’s collateral increases in value and enables them to secure funding.

The natural logarithm of the VIX index is a measure of forward uncertainty in the value of the firm's assets. Increasing the uncertainty will hinder the ability of the markets to assess the true probability of default. As a result, risk averse assessors will demand "excessive" 5-year spread, reducing the term premium of the CDX index.

Additional features in our model are the interrelation between the stochastic elements of each component and the endogenous shift of their volatility between regimes. This feature enables us to capture the occasional and recurrent endogenous regime switches of volatilities in time series. To understand the rationale for assuming regime shifts in the two components' disturbance terms consider the evidence on the standard deviations for different time periods of the term premium presented in Table VII-1⁴⁵. For the sample period between August 2004 and November 2007, the CDX term premium fluctuates narrowly around its mean at 23.007 with a modest standard deviation of 3.101. However, once interest rates began to rise and housing prices started to drop between 2006 and 2007 in many parts of the US, the refinancing of mortgages (especially the sub-prime mortgages) became extremely difficult. Defaults and foreclosure on those mortgages increased dramatically, which brought the sub-prime mortgage industry to the edge of collapse, and hence generated considerable uncertainty in financial markets. The standard deviation of the CDX term premium between November 2007 and July 2009 jumps to 18.804, which is more than 6 times higher than the aforementioned period between August 2004 and November 2007. Concurrently, the mean of the term premium falls to -16.913, suggesting that the CDX term premium might be experiencing a different regime in terms of both mean and volatility.

Traditionally, a sudden shift in the mean and volatility level of a time series would be modeled as a "structural break", in which this shift is due to some permanent changes in the economy's structure. Either can we pre-select the break points based on our prior or we let the data itself determine the break points endogenously (data-driven approach). However, the issue of identifying a structural break within a finite sample is a subtle one. A criticism of pre-selecting the break points is that this may lead to data-snooping. In addition, this method assumes these shifts in the economy's structure are deterministic and give no guidance on when they will appear again in the future. The data-driven approach of testing structural breaks also suffers a well-known criticism. A long time span of data is usually requested to

⁴⁵ Details of the data used in this paper are described in section VII.3.

obtain consistent parameters, yet structural break tests require these parameters to be estimated by splitting the finite sample into even smaller subsamples. The search of structural breaks over small subsamples, as argued in Lo and MacKinlay (1990), can bias inference toward finding breaks where none exist, especially in a very persistent covariance stationary time series. This last criticism may be particularly relevant to our study here, in which our sample period ends on the July 23rd, 2009, only a short time after the astonishing decline in the credit market yet still in the big uncertainty about future economy's recovery. Given the data available to us, an alternative method to model the recent abrupt changes in the CDX term premium is to assume the changes are recurrent. By allowing for regime switches in volatility (and in mean) to take place endogenously, we do not explicitly set a switching threshold value but we allow for the data to decide endogenously when to switch to a different regime.

Table VII-1: Mean and standard deviation of the CDX 5-year (CDX5Y), CDX 10-year (CDX10Y) and Term Premium (TP) for different sample periods

	Aug. 05, 2004 - Nov. 27, 2007			Nov.28, 2007 - Jul.23, 2009			Aug.05, 2004 - Jul.23, 2009		
	CDX5Y	CDX10Y	TP	CDX5Y	CDX10Y	TP	CDX5Y	CDX10Y	TP
Mean	46.789	69.795	23.007	159.322	142.409	-16.913	87.316	95.946	8.630
Std. Dev.	11.363	10.512	3.101	50.501	34.372	18.804	62.613	41.371	22.379

This regime switching attribute in the unobserved components space allows us to generate probabilities that each component of the term premium experiences either high or low volatility regimes through time. Although it complicates the estimation procedures - since additional filters must be employed to make inference on the hidden Markov chain process - allowing the two components to depend on different states of the economy provide us with an alternative approach to deal with the potential heteroskedastic variance in the daily CDX index series. The more conventional way of testing for financial time series heteroskedasticity is to consider ARCH-type volatility models, which allow constant unconditional volatility but time-varying conditional volatility. However, neglecting possible regime shifts in the unconditional variance, as shown in Lamoureux and Lastrapes (1990), would overestimate the persistence of the variance of a time series.

The remaining subsections present our stylized model of analysis. We first show how to construct the two components that drive the evolution of the CDX term premium. We

outline the state space representation of the system and our extension of modeling Markov switching disturbance terms.

VII.2.1 Stationary and Random Walk Components in State Space Representation

Let $X_{1,t}$ represents the stationary component that drives the term premium, and assume that $X_{1,t}$ is an Ornstein-Uhlenbeck process, whose dynamic evolution can be described by the stochastic differential equation

$$dX_{1,t} = k(\delta - X_{1,t})dt + \tilde{\sigma}_1 dZ_{1,t} \quad (\text{VII.1})$$

where δ is the target equilibrium or mean value supported by fundamentals; $\tilde{\sigma}_1 > 0$ is the scale of volatility that the exogenous shocks can transmit to the dynamics of $X_{1,t}$; $dZ_{1,t}$ is the standard Brownian motion with zero mean and unity variance that generate random exogenous shocks; $k > 0$ is the rate by which these shocks dissipate and the variable, $X_{1,t}$, reverts back to its mean. The Ornstein-Uhlenbeck process is an example of a Gaussian process that admits a stationary probability distribution and has a bounded variance. In contrast to the Brownian motion process that has constant drift term, the former allows for a drift term that is dependent on the current value of the process. If the current value of the process is lower than its long-term mean value, the drift term will be positive in order to bring the process back to its long-term mean value. If, on the other hand, the current value of the process is greater than its long-term mean value, the drift term will be negative in order to drag down the process back to its long-term mean value. In other words, this is a *mean-reversion process*. Setting $f(X_{1,t}, t) = X_{1,t}e^{kt}$ and applying the Ito's Lemma to this function, this leads to

$$df(X_{1,t}, t) = k\delta e^{kt}dt + \tilde{\sigma}_1 e^{kt}dZ_{1,t} \quad (\text{VII.2})$$

Integrating both sides of Equation (VII.2), we obtain

$$X_{1,t} = X_{1,0}e^{-kt} + \delta(1 - e^{-kt}) + \tilde{\sigma}_1 \int_0^t e^{-k(t-s)} dZ_{1,s}, \quad 0 \leq s \leq t \quad (\text{VII.3})$$

where $X_{1,0}$ is the initial value of the process and the first and the second moments are given by

$$\begin{aligned} E(X_{1,t}) &= X_{1,0}e^{-kt} + \delta(1 - e^{-kt}) \\ \text{Var}(X_{1,t}) &= \frac{\tilde{\sigma}_1^2(1 - e^{-2kt})}{2k} \end{aligned} \quad (\text{VII.4})$$

The Ornstein-Uhlenbeck process is one of several widely used approaches to model stochastically interest rates, exchanges rates and stock prices. The advantages of its simple and tractable solutions, under continuous-time framework, have been embodied in many empirical asset pricing models. The econometric modeling, however, emphasizes the discrete-time representation of stochastic processes. Consequently, the exact discrete time model correspond to Equation (VII.1) is given by the following AR(1) process

$$X_{1,t} = \delta(1 - e^{-k\Delta t}) + e^{-k\Delta t} X_{1,t-1} + \sigma_1 \Delta Z_{1,t} \quad (\text{VII.5})$$

where $\Delta t = 1/250$ is the sampling interval and $\sigma_1 = \tilde{\sigma}_1 \sqrt{\frac{(1 - e^{-k\Delta t})}{2k}}$. It is easy to see that $k > 0$ implies $e^{-k\Delta t} < 1$ and hence stationarity, $k \rightarrow 0$ or $\Delta t \rightarrow 0$ implies $e^{-k\Delta t} \rightarrow 1$ and the model converges to a unit root model.

Now, let $X_{2,t}$ be the second component that drives the term premium. We assume that it follows a driftless Random Walk (RW) process as shown in Equation (VII.6)

$$dX_{2,t} = \sigma_2 dZ_{2,t} \quad (\text{VII.6})$$

where σ_2 is the scaled volatility parameter and $dZ_{2,t}$ is the standard Brownian motion that can be assumed to be either dependent or independent of $dZ_{1,t}$. The discrete time version of Equation (VII.6) yields

$$X_{2,t} = X_{2,t-1} + \sigma_2 \Delta Z_{2,t} \quad (\text{VII.7})$$

The RW process has long been a popular choice for modeling the price dynamics of financial assets. In continuous time financial models, the price of stocks and stock indexes are modeled as geometric Brownian motions. It is relatively straightforward to show that the geometric Brownian motion of the price dynamic is equivalent to a RW path followed by the logarithm of the price in discrete time. The efficient market hypothesis in fact states that the financial asset's price follows a RW process, which literally assumes that the asset's price at time t is

determined by the price at the previous time period and the instantaneous price impact of the new flow of information. Although a RW process, like the one described in Equation (VII.7), has infinite unconditional mean and variance, the conditional mean and variance can be measured as

$$\begin{aligned} E_t(X_{2,t}) &= X_{2,t-1} \\ \text{Var}_t(X_{2,t}) &= \sigma_2^2 \end{aligned} \quad (\text{VII.8})$$

where the conditional expectation of the process at current time t depends only on the observation at previous time period.

Given the two unobserved components, constructed using Equation (VII.1) to Equation (VII.7), we estimate the parameter space as given by the system in Equation (VII.9), with the dynamics of the two components in a Bayesian updating manner, namely the Kalman filter algorithm based on a State Space system. State space representation is usually applied on dynamic time series models that involve unobserved variables (see, e.g., Engle and Watson (1981), Hamilton (1994b), Kim and Nelson (1989)). In our modeling, the fact that the two driving forces of the CDX index term premium - stationary and random walk components - are assumed to be unobserved state variables leads to the justification of using the state space representation. A typical state space model consists of two equations. One is a state equation that describes the dynamics of unobserved variables, which is shown below in Equation (VII.9); and the other one is the measurement equation that describes the relation between measured variables and the unobserved state variables, as shown in Equation (VII.10).

$$\begin{aligned} \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} &= \begin{bmatrix} \delta(1-e^{-k\Delta t}) \\ 0 \end{bmatrix} + \begin{bmatrix} e^{-k\Delta t} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \\ \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} &\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 \end{bmatrix} \Delta t\right) \end{aligned} \quad (\text{VII.9})$$

$$Y_t = X_{1,t} + X_{2,t} \quad (\text{VII.10})$$

In Equation (VII.9), the covariance terms $\sigma_1\sigma_2\rho_{12}$ and $\sigma_2\sigma_1\rho_{21}$ will be zero under the assumption of independence between the two disturbance terms (the correlation between the two disturbance terms - ρ_{12} - is zero).

In compact form, Equation (VII.9) can be rewritten as

$$\begin{aligned} X_t &= C + FX_{t-1} + \Sigma_t, \\ \Sigma_t &\sim N(0, Q) \end{aligned} \quad (\text{VII.11})$$

where $X_t = \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix}$, $C = \begin{bmatrix} \delta(1 - e^{-k\Delta t}) \\ 0 \end{bmatrix}$, $F = \begin{bmatrix} e^{-k\Delta t} & 0 \\ 0 & 1 \end{bmatrix}$, $\Sigma_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$ and

$$Q = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 \end{bmatrix} \Delta t.$$

The measurement equation, as described by Equation (VII.10), links linearly the term premium of the CDX index to the stationary and RW components. Rewriting this expression in a compact form, Equation (VII.10) reduces further to give

$$Y_t = HX_t \quad (\text{VII.12})$$

where Y_t is the term premium series and $H = [1 \ 1]$ represents the weights of two components in the term premium.

VII.2.2 State Space Model with Markov Switching Disturbances

An additional feature of our model is to allow each component's disturbance term to depend on different states of the economy. In practice, we let the volatilities of the disturbance terms to switch between high and low volatility regimes. Formally, we assume that σ_1^2 and σ_2^2 in Equation (VII.9) are time-varying and driven by two discrete-valued, independent unobserved first-order Markov chain processes $S_{1,t} = \{0,1\}$ and $S_{2,t} = \{0,1\}$ given by

$$\begin{aligned} \sigma_1^2 &= (1 - S_{1,t})\sigma_{1H}^2 + S_{1,t}\sigma_{1L}^2, \quad \sigma_{1H}^2 > \sigma_{1L}^2 \\ \sigma_2^2 &= (1 - S_{2,t})\sigma_{2H}^2 + S_{2,t}\sigma_{2L}^2, \quad \sigma_{2H}^2 > \sigma_{2L}^2 \end{aligned} \quad (\text{VII.13})$$

When both $S_{1,t}$ and $S_{2,t}$ are zeros, the two components will be in the high volatility state as $\sigma_1^2 = \sigma_{1H}^2$ and $\sigma_2^2 = \sigma_{2H}^2$; similarly if both $S_{1,t}$ and $S_{2,t}$ equal 1, the two components will be in the low volatility state since $\sigma_1^2 = \sigma_{1L}^2$ and $\sigma_2^2 = \sigma_{2L}^2$. The two remaining scenarios then categorize situations where the first component is in the high volatility state while the second is in the low ($S_{1,t} = 0, S_{2,t} = 1$) and where the first component is in the low volatility state while the second is in the high ($S_{1,t} = 1, S_{2,t} = 0$). In the simplest case of independent switching, which is when the discrete variables $S_{1,t}$ and $S_{2,t}$ involve independently of its own historical values, we can specify the transition probabilities vary simply as

$$\begin{aligned} p_{1,00} &= \Pr[S_{1,t} = 0] = p = \frac{\exp(p_0)}{1 + \exp(p_0)} \\ p_{1,11} &= \Pr[S_{1,t} = 1] = 1 - p = 1 - \frac{\exp(p_0)}{1 + \exp(p_0)} \\ p_{2,00} &= \Pr[S_{2,t} = 0] = q = \frac{\exp(q_0)}{1 + \exp(q_0)} \\ p_{2,11} &= \Pr[S_{2,t} = 1] = 1 - q = 1 - \frac{\exp(q_0)}{1 + \exp(q_0)} \end{aligned},$$

where p_0 and q_0 are the unrestricted parameters. In a more complicated case, where $S_{1,t}$ and $S_{2,t}$ follow Markov chain process, the evolution of $S_{1,t}$ and $S_{2,t}$ are dependent upon their historical values $S_{1,t-1}, S_{1,t-2}, \dots, S_{1,t-r}, S_{2,t-1}, S_{2,t-2}, \dots, S_{2,t-r}$, in which case the process of $S_{1,t}$ and $S_{2,t}$ is named as an r -th order Markov Switching process. In this paper, we consider the simplest first-order Markov switching process for $S_{1,t}$ and $S_{2,t}$, which means that the current value of the process at time t depends only on its previous value at time $t-1$. The likelihood for the process to remain at the previous value or change to the alternative depends on the transition probabilities from one state to the other, which are shown below as

$$\begin{aligned} p_{1,00} &= \Pr[S_{1,t} = 0 | S_{1,t-1} = 0] \\ p_{1,11} &= \Pr[S_{1,t} = 1 | S_{1,t-1} = 1] \\ p_{2,00} &= \Pr[S_{2,t} = 0 | S_{2,t-1} = 0] \\ p_{2,11} &= \Pr[S_{2,t} = 1 | S_{2,t-1} = 1] \end{aligned} \tag{VII.14}$$

Equivalently, the two transition probability matrices for each disturbance term can be written as:

$$p_1 = \begin{bmatrix} p_{1,00} & p_{1,01} \\ p_{1,10} & p_{1,11} \end{bmatrix}, p_2 = \begin{bmatrix} p_{2,00} & p_{2,01} \\ p_{2,10} & p_{2,11} \end{bmatrix} \quad (\text{VII.15})$$

where $p_{q,ij} = \Pr[S_{q,t} = j | S_{q,t-1} = i]$ with $\sum_{j=1}^2 p_{ij} = 1$, $\forall i$ and $q = \{1, 2\}$.

To estimation of the transition probabilities as shown above, it requires the choice of the appropriate functional forms of the probability functions that govern the Markov chain variables. Since the transition probabilities have to be bounded within $[0,1]$ the usual choice is the adoption of the logistic transformation on the probability terms as

$$\begin{aligned} p_{1,00} &= \Pr[S_{1,t} = 0 | S_{1,t-1} = 0] = \frac{\exp(d_{1,0})}{1 + \exp(d_{1,0})} \\ p_{1,01} &= 1 - p_{1,00} \\ p_{1,11} &= \Pr[S_{1,t} = 1 | S_{1,t-1} = 1] = \frac{\exp(d_{1,1})}{1 + \exp(d_{1,1})} \\ p_{1,10} &= 1 - p_{1,11} \\ p_{2,00} &= \Pr[S_{2,t} = 0 | S_{2,t-1} = 0] = \frac{\exp(d_{2,0})}{1 + \exp(d_{2,0})} \\ p_{2,01} &= 1 - p_{2,00} \\ p_{2,11} &= \Pr[S_{2,t} = 1 | S_{2,t-1} = 1] = \frac{\exp(d_{2,1})}{1 + \exp(d_{2,1})} \\ p_{2,10} &= 1 - p_{2,11} \end{aligned} \quad (\text{VII.16})$$

where $d_{1,0}$, $d_{1,1}$, $d_{2,0}$ and $d_{2,1}$ are the unconstrained parameter.

VII.2.3 Estimation procedure

To estimate the state space Markov switching model, described in previous subsections, we use Kim's filter (Kim (1994)), which is a numerical algorithm that combine the Kalman filter in estimating state space models and the Hamilton filter (Hamilton (1989a)) in estimating Markov switching models. In the conventional derivation of the Kalman filter for an invariant parameter state space model, the goal is to make predictions of the unobserved state variables based on the current information set, denoted $X_{t|t-1} = E(X_t | I_{t-1})$, where I_{t-1} represents all observed variables available at time $t-1$. The mean squared error of

the prediction, denoted as $P_{t|t-1}$, is $P_{t|t-1} = E\left((X_t - X_{t|t-1})(X_t - X_{t|t-1})' | I_{t-1}\right)$. The Kalman filter algorithm then implements a sequence of Bayesian updating on the unobserved variable X_t and the mean square error P_t when observing a new data entry. The updated unobserved variable $X_{t|t}$, given the observation of information the set at time t , is formed as a weighted average of $X_{t|t-1}$ and new information contained in the prediction error, where the weight assigned to this new information is called Kalman gain. This prediction and updating process evolve over time and are conditional on the correctly estimated parameters of the model. As a result, Kalman filter will need to be initialized in the first place. Specifically, the initial value of X_t and its mean square error at time 0 conditional on the information up to time 0 ($X_{0|0}$ and $P_{0|0}$) need to be assigned with some carefully chosen initial values. For stationary X_t , $X_{0|0}$ and $P_{0|0}$ can be assigned with the unconditional mean and covariance matrix of X_t . For non-stationary X_t (or partially non-stationary X_t as in our case), however, the unconditional mean and covariance matrix of X_t do not exist. In this case, we follow Kim (1994), Kim, Piger and Startz (2007), Morley and Piger (2008) to arbitrarily set $X_{0|0}$ at some values based on wild guessing, and subsequently to tackle this very large uncertainty due to the wild guesses by assigning some very large values to the diagonal of $P_{0|0}$. Finally, the prediction errors and their variances, as the by-products of the prediction process, will be used to construct the log-likelihood function

$$L(\theta) = -\frac{1}{2} \sum_{t=\tau+1}^T \ln \left[(2\pi)^n |\omega_{t|t-1}| \right] - \frac{1}{2} \sum_{t=\tau+1}^T \psi'_{t|t-1} \omega_{t|t-1}^{-1} \psi_{t|t-1} \quad (\text{VII.17})$$

where $\psi_{t|t-1}$ is the prediction error and $\omega_{t|t-1}$ is its conditional variance. As noticed, this log-likelihood value function will be evaluated from $t = \tau + 1$, where τ is set to be large enough (in our case, we set $\tau = 10$) in order to minimize the effect of the arbitrary initial value $X_{0|0}$ on the log-likelihood value.

In addition to the nonstandard problem of initializing non-stationary state factors in a state space system, the inferential procedures on the Markov switching variables ($S_{1,t}$ and $S_{2,t}$) would undoubtedly complicate the estimation procedures. The prediction and updating processes of the unobserved variable X_t now depend on both the previous and current values

of the Markov variables. Since we have two independent Markov chain processes in our model, for given realizations of the two Markov variables at times t and $t-1$ ($S_{1,t-1} = i$, $S_{1,t} = j$, $S_{2,t-1} = i$ and $S_{2,t} = j$, where $i = \{0,1\}$, $j = \{0,1\}$), the Kalman filter equations can then be represented as follows

$$\begin{aligned}
X_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} &= C + FX_{t-1|t-1}^{S_{1,t-1}S_{2,t-1}} \\
P_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} &= FX_{t-1|t-1}^{S_{1,t-1}S_{2,t-1}} F' + \Sigma^{S_{1,t}S_{2,t}} \\
\eta_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} &= Y_t - HX_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} \\
f_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} &= HP_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} H' \\
X_{t|t}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} &= X_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} + \frac{P_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} H'}{f_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}} \eta_{t|t-1}^{S_{1,t-1}S_{2,t-1}} \\
P_{t|t}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} &= P_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} - \frac{P_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} H' HP_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}}{f_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}}
\end{aligned} \tag{VII.18}$$

where $X_{t-1|t-1}^{S_{1,t-1}S_{2,t-1}}$ is the value of X_{t-1} based on the information up to time $t-1$, given that $S_{1,t-1} = i$ and $S_{2,t-1} = i$; $X_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$ is the updated value of X_t based on the information up to time $t-1$, given that $S_{1,t-1} = i$, $S_{1,t} = j$, $S_{2,t-1} = i$ and $S_{2,t} = j$; $P_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$ is the mean squared error of the unobserved $X_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$ given $S_{1,t-1} = i$, $S_{1,t} = j$, $S_{2,t-1} = i$ and $S_{2,t} = j$; $\eta_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$ is the prediction error of Y_t in the measurement equation, given the updated forecast of X_t as $X_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$ conditional on $S_{1,t-1} = i$, $S_{1,t} = j$, $S_{2,t-1} = i$ and $S_{2,t} = j$ based on the information up to time $t-1$; $f_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$ is the conditional variance of the forecast error $\eta_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$; $X_{t|t}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$ and $P_{t|t}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$ are the updated X_t and P_t based on the information up to time t , given that $S_{1,t-1} = i$, $S_{1,t} = j$, $S_{2,t-1} = i$ and $S_{2,t} = j$.

Since each iteration of the Kalman filter produces a 4-fold increase in the number of cases to consider⁴⁶, we reduce the 16 one-period posteriors $X_{t|t}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$ and $P_{t|t}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$ into

⁴⁶ We have 4 cases to consider in each iteration of the Kalman filter: (1) both stationary and random walk components are in high volatility regime; (2) stationary component is in the high volatility regime while the random walk component is in the low volatility regime; (3) stationary component is in the low volatility regime while the random walk component is in the high volatility regime; (4) both stationary and random walk components are in low volatility regime. Therefore, every new iteration, the first order dependence of the current Markov chain variable on its previous value leads to a 4-fold increase in the number of cases to consider.

4 by taking appropriate approximations at the end of each iteration. This is computed through Kim's approximation procedures

$$X_{t|t}^{S_{1,t}S_{2,t}} = \frac{\sum_{S_{1,t-1}=0}^1 \sum_{S_{2,t-1}=0}^1 \Pr(S_{1,t-1}=i, S_{1,t}=j, S_{2,t-1}=i, S_{2,t}=j | I_t) X_{t|t}^{S_{1,t}S_{1,t-1}S_{2,t}S_{2,t-1}}}{\Pr(S_{1,t}=j, S_{2,t}=j | I_t)} \quad (\text{VII.19})$$

$$\begin{aligned} P_{t|t}^{S_{1,t}S_{2,t}} &= \sum_{S_{1,t}=0}^1 \sum_{S_{2,t}=0}^1 \left(P_{t|t}^{S_{1,t}S_{1,t-1}S_{2,t}S_{2,t-1}} + \left(X_{t|t}^{S_{1,t}S_{2,t}} - X_{t|t}^{S_{1,t}S_{1,t-1}S_{2,t}S_{2,t-1}} \right) \left(X_{t|t}^{S_{1,t}S_{2,t}} - X_{t|t}^{S_{1,t}S_{1,t-1}S_{2,t}S_{2,t-1}} \right) \right) \\ &\times \frac{\Pr(S_{1,t-1}=i, S_{1,t}=j, S_{2,t-1}=i, S_{2,t}=j | I_t)}{\Pr(S_{1,t}=j, S_{2,t}=j | I_t)} \end{aligned} \quad (\text{VII.20})$$

where the probability terms in the above two equations are obtained from Hamilton's filter as

$$\begin{aligned} &\Pr(S_{1,t-1}=i, S_{1,t}=j, S_{2,t-1}=i, S_{2,t}=j | I_t) \\ &= \frac{\Pr(Y_t | S_{1,t-1}=i, S_{1,t}=j, S_{2,t-1}=i, S_{2,t}=j, I_{t-1}) \Pr(S_{1,t-1}=i, S_{1,t}=j, S_{2,t-1}=i, S_{2,t}=j | I_{t-1})}{\Pr(Y_t | I_{t-1})} \end{aligned}$$

with

$$\begin{aligned} &\Pr(Y_t | S_{1,t-1}=i, S_{1,t}=j, S_{2,t-1}=i, S_{2,t}=j, I_{t-1}) \\ &= \frac{1}{\sqrt{(2\pi)^N |f_{t|t-1}^{S_{1,t}S_{1,t-1}, S_{2,t}S_{2,t-1}}|}} \exp \left(-\frac{1}{2} \frac{(\eta_{t|t-1}^{S_{1,t-1}S_{2,t-1}})'(\eta_{t|t-1}^{S_{1,t-1}S_{2,t-1}})}{f_{t|t-1}^{S_{1,t}S_{1,t-1}, S_{2,t}S_{2,t-1}}} \right), \end{aligned}$$

$$\Pr(Y_t | I_{t-1}) = \sum_{S_{1,t-1}=0}^1 \sum_{S_{1,t}=0}^1 \sum_{S_{2,t-1}=0}^1 \sum_{S_{2,t}=0}^1 \Pr(Y_t, S_{1,t-1}=i, S_{1,t}=j, S_{2,t-1}=i, S_{2,t}=j | I_{t-1})$$

and

$$\begin{aligned} &\Pr(S_{1,t-1}=i, S_{1,t}=j, S_{2,t-1}=i, S_{2,t}=j | I_{t-1}) \\ &= \Pr(S_{1,t}=j | S_{1,t-1}=i) \Pr(S_{2,t}=j | S_{2,t-1}=i) \Pr(S_{1,t-1}=i, S_{2,t-1}=i | I_{t-1}) \end{aligned}$$

with

$$\Pr(S_{1,t-1} = i, S_{2,t-1} = i | I_{t-1}) = \sum_{S_{1,t-2}=0}^1 \sum_{S_{2,t-2}=0}^1 \Pr(S_{1,t-2} = i, S_{1,t-1} = i, S_{2,t-2} = i, S_{2,t-1} = i | I_{t-1})$$

At the end of each iteration, Equation (VII.19) and Equation (VII.20) are used to collapse 16 one-period posteriors ($X_{t|t}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$ and $P_{t|t}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}}$) into 4 ($X_{t|t}^{S_{1,t}S_{2,t}}$ and $P_{t|t}^{S_{1,t}S_{2,t}}$). As a by-product of the Hamilton filter, the approximate log likelihood function is given by

$$L(\theta) = \sum_{t=\tau+1}^T \ln f(Y_t | I_{t-1})$$

that will be maximized with respect to the parameter vector space $\Theta = \{p_{1,00}, p_{1,11}, p_{2,00}, p_{2,11}, \delta, k, a, \sigma_{1,H}, \sigma_{1,L}, \sigma_{2,H}, \sigma_{2,L}, \rho_{12}\}$.

VII.3 Data

CDS contracts are by nature over-the-counter. In this market, an interested party searches through brokers or dealers to find counter-parties. The terms of a contract is then negotiated between the two parties and the information of trade is passed to a clearing house. Because the entire process is not sufficiently standardised, the availability and quality of the data are not as dependable as those of exchange-based transactions. In general, the CDS market is relatively illiquid compared to other established markets, as noted in Tang and Yan (2007), the bid-ask spread is high – at 23% on average – with a sizable fixed component. Yet, the development of the CDS market is tremendous over the last decade, the notional amount of single-name CDS, baskets and portfolio of credits and index trades, as reported in The International Swaps and Derivatives Association (ISDA) 2006 Year-End Market Survey, reached \$34.4 trillion by 2006. Single-name CDS are the most liquid among many credit derivatives traded in market. The figures from the Bank of International Settlements (BIS) show that among the notional amount of \$28.8 trillion for credit derivatives by the end of 2006, \$18.9 trillion is for single-name CDS contracts. Accordingly to the 2003/04 Credit Derivatives Reports by the British Bankers' Association (BBA), the typical maturity of a CDS contract is 5 years. The typical notional amount is \$5-10 million for investment grade credits and \$2-5 million for high-yield credits. London and New York each accounts for about 40% of the total CDS market with 86% of the transactions use physical settlement.

CDS data are mainly collected by large investment banks which only record their own entering transactions. Although professional data vendors are the primary source of CDS

research data, the data suffer the following pitfalls: (1) as in Zhu (2006), data prior to 1999 are very limited; (2) the frequency of CDS transaction data is low, therefore, the usual instantaneous lead-lag analysis which use much higher frequency data may not be robust; (3) CDS prices are usually obtained as “quoted” prices which may not reflect the actual information contained in trading prices; (4) CDS data are truncated in the way that the majority contracts have a maturity of 5 years with nominal amount of \$5 million or \$10 million; (5) As noted in Blanco, et al. (2005), CDS data are usually unevenly spaced with missing or, very occasionally, suspicious observations in time series. To overcome the above limitations, previous researches (for example, Blanco, et al. (2005), Cossin, Hricko, Aunon-Nerin and Huang (2002)) use multiple data sources including CreditTrade, a CDS broker, and J.P.Morgan. The missing values are filled with J.P. Morgan’s mid-market prices since they are rarely outside the bid-ask spread quotes from CreditTrade. Suspicious entries are checked for general confirmation of the CreditTrade data.⁴⁷

To circumvent the above mentioned limitations on Single-name CDS time series, we restrict our sample of analysis to the CDS tranche index market, specifically to the North American CDX investment-grade indices. In these indices, all 125 single-name credits have equal weights in the portfolio. The Dow Jones CDX IG five-year index is a basket of CDSs on 125 names for the U.S. investment-grade market. Each reference entity has a weight of 0.8%. We use a representative dataset of daily CDS prices for the North American CDX tranche index and focus our analysis on the most liquid segments of the CDX index market, which are the 5-year and 10-year maturities. The analysis is based on daily data spanning from 2004 to 2009. The main source of CDX data is Markit.

In our sample period (August 5, 2004 to July 23, 2009), as shown in Table VII-2 and Figure VII-1, the average premium is 87.3161 basis points for the CDX 5-year index and 95.9463 for the CDX 10-year index. The CDX 5-year index reaches its maximum spread (283 basis points) on September 16, 2008, which is the day after the announcement of Lehman Brothers default. Similarly, the CDX 10-year index reaches its maximum premium (251 basis points) on the same day, as it is reflected in the negative slope of the credit curve. Not only the volatility of the CDS indices increased dramatically as of July 2007, but also the levels of these indices started to rise significantly at the onset of the crisis. This

⁴⁷ An in-depth discussion of CDS data quality can be found in Blanco, *et al.* (2005).

unprecedented financial turmoil directly raised investors' expectations on imminent future defaults of firms' debts, especially of those financial firms heavily exposed to sub-prime-mortgage lending. From the second panel of Figure VII-1, we can see that the term premium of the CDX 10-year – 5-year index becomes negative at the start of 2008, suggesting an increase of investors' concerns over short-term default risk. Nevertheless, market sentiment remains unchanged over long-term horizons.

Daily Federal Fund Rates (*FFR*) are obtained from the St. Louis FRED database. Daily slope of the yield curve (*SLOPE101*), calculated using 10-year and 1-year U.S. Treasury bond yields, are from Thomson Reuters© Datastream. S&P 500 index daily return (*SP500RTN*) observations are obtained from Thomson Reuters© Datastream. VIX index (*VIX*) data are from the Chicago Options Mercantile Exchange (CBOE). The descriptive statistics of monetary policy and equity market condition variables are also presented in Table VII-2 and Figure VII-2 depicts their evolution over the sample period.

Table VII-2: Descriptive statistics of the data series

	CDX5Y	CDX10Y	TP	FFR	SLOPE101	SP500RT N	VIX
Mean	87.316	95.946	8.630	3.289	1.268	-0.0001	21.236
Median	53.000	77.000	22.000	3.620	1.045	0.0007	15.630
Maximum	283.371	251.363	33.000	5.410	4.010	0.104	80.860
Minimum	29.000	54.000	-53.333	0.080	-0.780	-0.095	9.890
Std. Dev.	62.613	41.371	22.379	1.811	1.233	0.016	12.877
Skewness	1.197	1.191	-1.332	-0.492	0.278	-0.253	1.898
Kurtosis	3.202	3.417	3.607	1.885	1.912	12.357	6.543
Jarque-Bera	281.280	284.875	363.461	105.915	71.541	4277.339	1312.913
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Observations	1169	1169	1169	1169	1169	1169	1169

Note: Jarque-Bera test is the test for the null of normality, p-values are reported in square brackets.

Figure VII-1: CDX-5 year, CDX-10 Year and CDX term premium

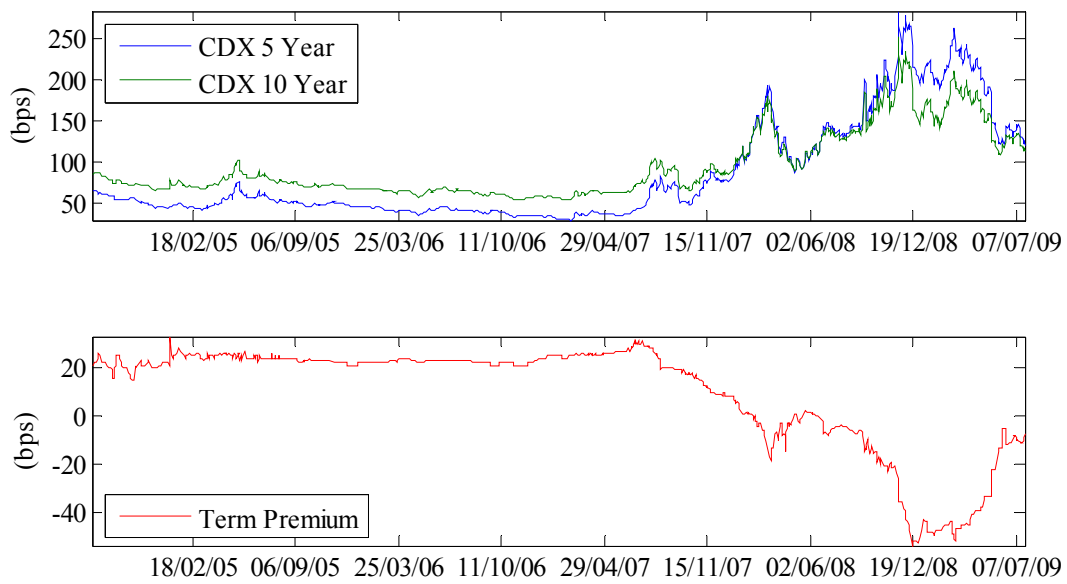
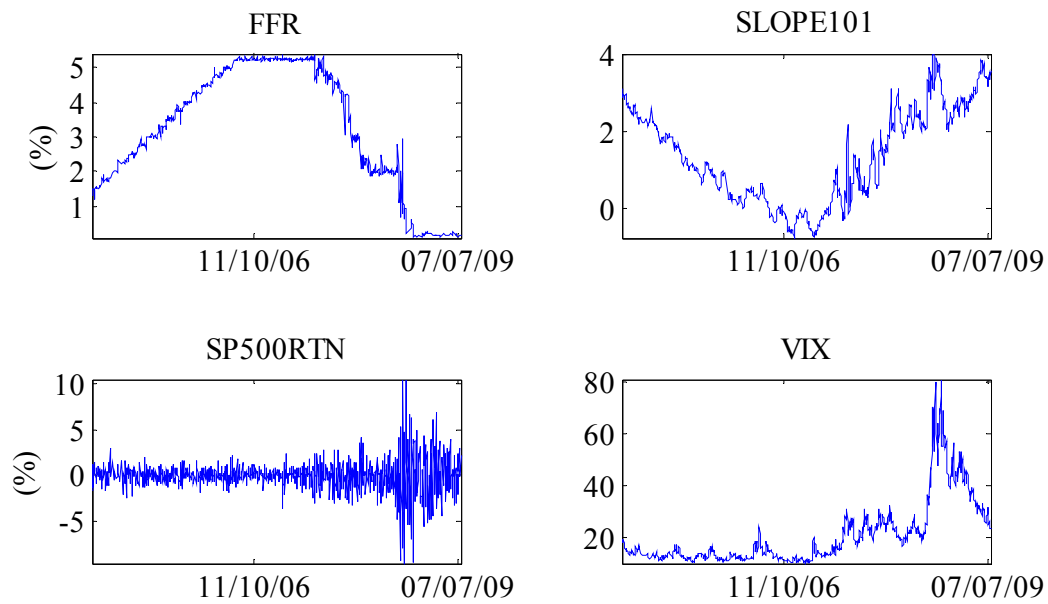


Figure VII-2: Actual plots of: *FFR*, *SLOPE101*, *SP500RTN* and *VIX*



VII.4 Empirical Results

Using the methodology described in section VII.2, we estimate a series of nested Markov-switching unobserved component models, followed by a battery of tests on model specification to determine the preferred model that will be used in the empirical analysis.

VII.4.1 Model selection tests

It is well known that for Markov-switching models the standard likelihood ratio test of the null hypothesis of linearity does not have the usual χ^2 distribution. The reason is that there are nuisance parameters which cannot be identified under the null hypothesis. As a result, the scores evaluated at the null hypothesis are identically zero. Hansen (1992) and Garcia (1998) introduce alternative tests of the linearity against regime switching. In this paper, we use Hansen (1992) procedure, which provides an upper bound of the p – value for linearity, to determine the significance of improvement of allowing Markov-switching disturbance terms in the two components. In addition, we also consider more conventional ways of selecting models based on the Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (BIC). Finally, we verify our model selection results by running a series of residual diagnostic tests to see if the selected model could capture serial correlation and heteroskedasticity in the data series.

To implement Hansen (1992)'s procedure, we need to evaluate the constrained likelihood under the null hypothesis over a grid of values for the nuisance parameters. As noted in Hansen (1992), the only practical way to evaluate the maximal statistics is to form a grid search over a relatively small number of values of the nuisance parameters. A trade-off arises since a more extensive grid search means a major computational burden, but reduces the arbitrariness associated with the choice of grid. Defining the restricted model under null hypothesis of no regime switching of the two components' disturbance terms as consisting of Equations 9-10 with $\rho_{12} = \rho_{21} = 0$, and the alternative model under assumption of Markov-switching disturbance terms as involving Equations (VII.13)-(VII.16), we have nuisance parameters $\{\sigma_{1H}, \sigma_{2H}, p_{1,00}, p_{1,11}, p_{2,00}, p_{2,11}\}$. To select a proper grid for these nuisance parameters, we follow Kim, Morley and Piger (2005) to consider a grid covering the likely values of these when estimating the Markov switching alternative. The grids that we used for σ_{1H} and σ_{2H} are $[5, 10.5]$ and $[1.5, 5]$, respectively, each in increment step of 0.5. The grids for $\{p_{1,00}, p_{2,00}\}$ varies from 0.7 to 0.9 in increment step of 0.05 and for $\{p_{1,11}, p_{2,11}\}$ from 0.8 to 0.9 in increment step of 0.05. The Hansen test applied as described above, yields a conservative p – value of 0.001, which clearly indicate a strong rejection of linearity in favoring of Markov-switching disturbance terms.

To determine whether a model permits correlated disturbance terms better performs a model restricts zeros correlations, we look at Table VII-3 and Table VII-4. While AIC and BIC select correlated models, likelihood ratio tests show Model 1 and Model 2 better fit the data than Model 5 and 6, respectively. However, the improvement in likelihood value when allowing correlations and all parameters to switch between regimes (Model 8) is overwhelming. Meanwhile, we also report the likelihood ratio tests within each nested group in Table VII-5, from which we can see the most flexible model (Model 8) performs the best. We verify this result with the residual diagnostic tests in Table VII-6, where we test the overall randomness of the residuals of the models (summation of the disturbance terms of the two components) with the null hypothesis of assuming randomness. We report two Ljung-Box Q statistics for each model: one is the autocorrelation Q statistics based on the standardized residuals up to 20 lags; the other one is the ARCH effect Q statistics based on the squared standardized residuals up to 20 lags. Looking at Table VII-6, the Q statistics suggest Model 8 as the best performing model, since it captures both the autocorrelation and ARCH effects in the residuals.

Table VII-3: Model selection results

Model Specifications	No. of parameters	AIC	BIC	$\ln L$
Model 0: $\rho_{12} = \rho_{21} = 0$ (single regime)	4	3.135	3.152	-1798.390
Model 1: $\rho_{12} = \rho_{21} = 0$	10	2.800	2.844	-1600.010
Model 2: $\rho_{12} = \rho_{21} = 0$ (k is regime dependent)	11	2.777	2.826	-1585.916
Model 3: $\rho_{12} = \rho_{21} = 0$ (δ is regime dependent)	11	2.763	2.811	-1577.435
Model 4: $\rho_{12} = \rho_{21} = 0$ (both k and δ are regime dependent)	12	2.690	2.743	-1534.727
Model 5: $\rho_{12} = \rho_{21} \neq 0$	14	2.801	2.863	-1596.695
Model 6: $\rho_{12} = \rho_{21} \neq 0$ (k is regime dependent)	15	2.782	2.848	-1584.757
Model 7: $\rho_{12} = \rho_{21} \neq 0$ (δ is regime dependent)	15	2.762	2.828	-1573.281
Model 8: $\rho_{12} = \rho_{21} \neq 0$ (both k and δ are regime dependent)	16	2.671	2.741	-1519.815

Note: Model 0 refers to system of Equations VI.9-VI.10 assuming $\rho_{12} = \rho_{21} = 0$. Model 1 builds on Model 0 with Markov-switching variances defined in Equations 13-16; Model 2 builds on Model 1 but allow k to switch regimes; Model 3 builds on Model 1 but allow δ to switch regimes; Model 4 builds on Model 1 but allow both

k and δ to switch regimes; Models 5-8 differ from 1-4 for allowing correlations between the two components' disturbance terms. $\ln L$ denotes the natural logarithm of likelihood value. AIC denotes the Akaike Information Criterion and BIC denotes the Schwarz Bayesian Information Criterion. A smaller statistic of AIC or BIC corresponds to smaller estimated Kullback-Leibler distance from the true model.

Table VII-4: Likelihood ratio tests on constraint $\rho_{12} = \rho_{21} = 0$

Constraint: $\rho_{12} = \rho_{21} = 0$	Likelihood ratio	p-value
Model 1 to Model 5	6.629	0.157
Model 2 to Model 6	2.318	0.678
Model 3 to Model 7	8.309	0.081
Model 4 to Model 8	29.825	5.313E-06

Table VII-5: Likelihood ratio tests within groups

Constraints	Likelihood ratio	p-value
Group of models apply $\rho_{12} = \rho_{21} = 0$		
Model 1 to Model 2 (k is regime dependent)	28.187	0.000
Model 1 to Model 3 (δ is regime dependent)	45.149	1.826E-11
Model 1 to Model 4 (both k and δ are regime dependent)	130.565	4.447E-29
Group of models apply $\rho_{12} = \rho_{21} \neq 0$		
Model 5 to Model 6 (k is regime dependent)	23.876	1.027E-06
Model 5 to Model 7 (δ is regime dependent)	46.829	7.746E-12
Model 5 to Model 8 (both k and δ are regime dependent)	153.761	4.085E-34

Table VII-6: Residual diagnostic tests

Lags	Autocorrelation		ARCH		Autocorrelation		ARCH	
	Q-stats	p-value	Q-stats	p-value	Q-stats	p-value	Q-stats	p-value
	Model 0							
1	3.3651	0.0666	55.4375	1.00E-13				
5	6.4061	0.2687	56.5142	6.37E-11				
10	21.4859	0.0179	84.9913	5.00E-14				

20	51.2066	0.0001	128.0163	0.00E+00				
	Model 1				Model 5			
1	5.8659	0.0154	12.4945	0.0004	0.002	0.9643	1.4605	0.2268
5	38.7978	0	77.4846	0	21.916	0.0005	12.6867	0.0265
10	46.7586	0	98.3647	0	39.9475	0	21.0755	0.0206
20	127.264	0	157.386	0	91.3123	0	42.9094	0.0021
	Model 2				Model 6			
1	10.665	0.0011	2.0368	0.1535	12.7731	3.52E-04	2.2811	0.131
5	64.4022	0	6.0608	0.3003	69.3635	0.00E+00	5.6359	0.3433
10	129.4365	0	38.6211	0	140.9098	0.00E+00	48.6666	0
20	244.9473	0	59.528	0	256.8626	0.00E+00	67.5147	0
	Model 3				Model 7			
1	34.5333	4.19E-09	0.3524	0.5528	14.6746	1.28E-04	0.0017	0.9673
5	73.5529	0.00E+00	13.7145	0.0175	22.747	3.77E-04	0.0029	1
10	94.9219	0.00E+00	37.0741	0.0001	50.5479	2.00E-07	1.4622	0.999
20	114.7697	0.00E+00	97.5292	0	76.9609	0.00E+00	1.4862	1
	Model 4				Model 8			
1	9.6918	0.0019	0.0337	0.8544	1.0875	0.297	0.0044	0.9468
5	19.6283	0.0015	1.3234	0.9325	4.0601	0.5408	0.0107	1
10	37.2164	0.0001	13.1973	0.2128	9.7284	0.4646	0.0267	1
20	56.5368	0	16.2277	0.7024	22.5042	0.3138	0.1733	1

VII.4.2 Estimates of the Markov-switching unobserved component model

Table VII-7 reports the maximum likelihood estimates of Model 8, the most flexible and best performing model suggested by the model selection procedures considered above. The first noticeable result is the two regime dependent long term equilibriums of the stationary component: 8.4516 as in the low volatility regime and -0.5633 in the high regime. As can be seen, during the calm period (from August 05, 2004 to December 31, 2007), the slope along the credit curve is positive since the term premium is the compensation for default risk in 5 years time. However, since the outbreak of the crisis (August 2007), severe strains in financial markets, banks' assets writedowns and diminishing liquidity in funding

markets raised the level of uncertainty about corporate default risk. The inversion of the credit curve, as embedded in a negative CDX index term premium, vividly capture this deteriorating outlook.

Table VII-7: Estimation results of Model 8

Parameters	Estimate	Std. Error
δ_L	8.45161	0.19273
δ_H	-0.56331	0.22730
k_L	465.078	30.5714
k_H	23.9443	3.79152
$\sigma_{1,L}$	0.00656	0.01754
$\sigma_{1,H}$	7.23834	0.74797
$\sigma_{2,L}$	0.00071	0.00672
$\sigma_{2,H}$	4.11551	0.23609
$\rho_{1L,2L}$	0.88234	2.69776
$\rho_{1H,2L}$	-0.88688	23.16079
$\rho_{1L,2H}$	-0.62831	4.08741
$\rho_{1H,2H}$	-0.72941	0.21734
$p_{1,LL} (p_{1,00})$	0.99561	0.00149
$p_{1,HH} (p_{1,11})$	0.99999	1.68E-06
$p_{2,LL} (p_{2,00})$	0.96427	0.00110
$p_{2,HH} (p_{2,11})$	0.97346	0.00068
$\ln L$	-1519.81491	

Note: Subscript L denotes Low volatility regime and H denotes High volatility regime. The standard errors are numerically calculated using the variance-covariance matrix of the estimators (see e.g. Hamilton (1994, p.389) for details).

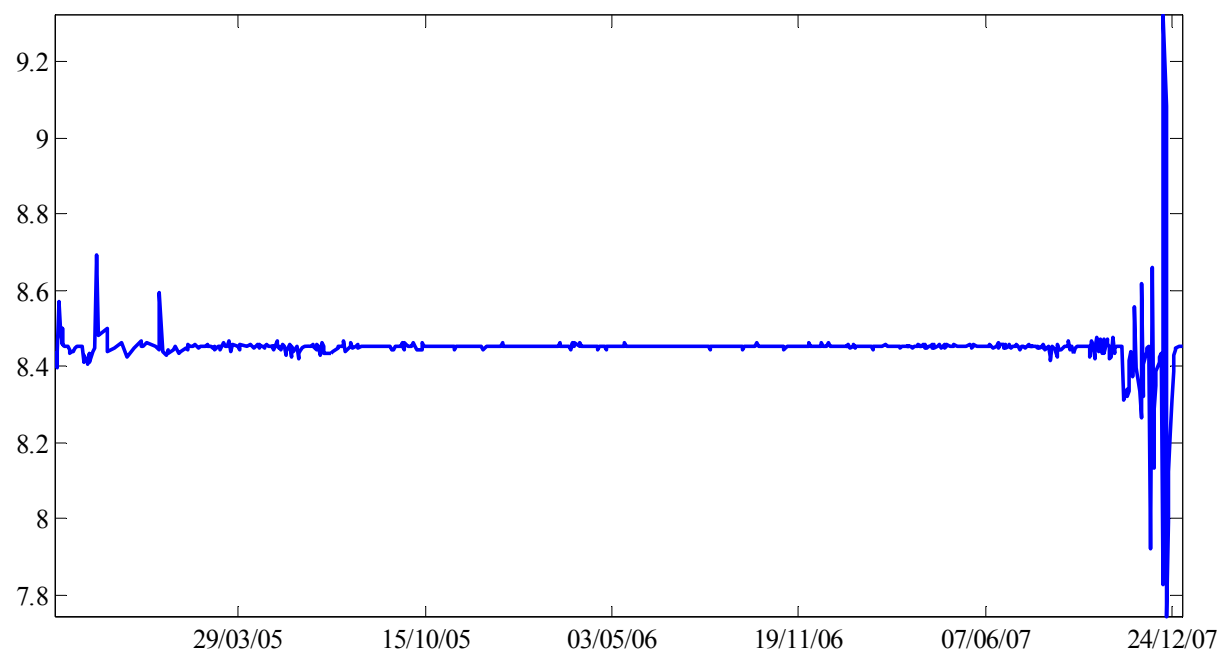
The second remarkable result is the regime dependent mean reverting speed for the stationary component. During non-crisis periods, asset prices are less likely to stay high or low period-to-period but mean revert quickly to their long-term equilibrium values. In other words, mean reverting assets prices imply a low probability of ending up in the tail of the distribution. Our estimation of the measure of mean reverting speed (k) is 465.07 in a low volatility regime, which transfers to a first-order autocorrelation of 0.1556. The speed in the high volatility regime, on the other hand, falls to 23.94 or 0.9087 in terms of first-order

autocorrelation, which suggests a very persistent behavior of the stationary component in the high volatility regime.

Next, we look at the volatilities of stationary and RW components in each regime. In terms of stationary shocks, the estimated volatility in the low volatility regime is positive but statistically insignificant. This is primarily due to the very high mean reverting speed in the low volatility regime. Therefore, the variations in the low volatility regime are barely zero. It is more clearly shown in Figure VII-3 that the stationary component's variation in low volatility regime is tight at around 8.45 and the largest change is only 1.4 basis points. On the other hand, the scale of the variation for the stationary component in the high volatility regime is by far higher than in the non-crisis period, with an estimated annual volatility of 7.238. One may notice the estimated standard errors are large for correlations between the variances of each component. This reflects the high degrees of estimation uncertainty around the correlation estimates (either positive or negative).

high correlation between the variances of stationary and random walk component commenced from 2008

Figure VII-3: Snapshot of the stationary component in its low volatility regime



The distinguishing feature of our model is that it allows us to decompose the term premium into two correlated driving components. The filtered RW and the stationary

components with the associated filtered state probabilities are displayed in Figure VII-4 and Figure VII-5 (see Appendix). In the first part of the sample period, temporary volatile movements in the RW and stationary components are induced by severe shocks such as the GM and Ford downgrade events of May 2005. In the subsequent period, the credit market enjoyed a rapid growth in terms of both trading volumes and products innovation. During this time period, both components stay in the low volatility regime, as reflected by a flat CDX credit curve. The frequent regime changes of the RW component start taking place in the aftermath of Countrywide's bankruptcy. In fact, on August 15, 2007, Countrywide Financial, the largest mortgage lender in the United States, announced that the foreclosure and mortgage delinquencies had risen to their highest level since early 2002. Since then, because of that episode and of the events around the onset of the subprime mortgage crisis, the CDX index term premium exhibits a downward trend.

On the other hand, as shown in Figure VII-4, the stationary component enters a high volatility period from the beginning of 2008. Increased credit-related writedowns for individual banks (e.g. Citigroup) owe to a further deterioration in the corporate debt and prime residential mortgage markets, as the crisis originally centered in subprime mortgages spilled over to adversely affect economic prospects more broadly. On January 22, 2008, the Federal Reserve cut the Fed funds target rate to 3.5% - an unprecedented decision, taken between scheduled meetings, and the largest single cut in 23 years. The move followed the biggest one day loss on world stock exchanges in almost six years. The rate is reduced further to 3% on January 30, 2008. On March 14, 2008, Bear Sterns' demise brought about a dramatic increase in stock market volatility and liquidity shortages in funding markets. As illustrated in Figure VII-4, the resulting decline of the stationary component on that day captures investor confidence-induced downward spirals. The subsequent abrupt jumps occur on July 13, 2008 and November 23, 2008 when the US authorities announced the nationalization of Fannie Mae and Freddie Mac and a rescue package of Citigroup.

Although the filtered state probability of the high volatility regime for stationary component accurately signal greater market volatility since early 2008, it is quite evident from inspection of this data that the filtered state probability of the high volatility regime for the RW component is close to zero at the occurrence of extreme credit events, such as the Bear Stearn's and Lehman Brothers' default announcements. At first sight, this counter-intuitive result may be difficult to understand. It conveys the message that credit market

uncertainties, as measured by the conditional variance of the term premium, decline significantly with the abrupt unveiling of tail risk events.

Figure VII-6 plots the conditional variance derived from Model 8. Using the definition of Kalman filter provided in the previous section, the conditional forecast error variance is given by

$$f_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} = HP_{t|t-1}^{S_{1,t-1}S_{1,t}S_{2,t-1}S_{2,t}} H' \quad (\text{VII.21})$$

which can be regarded as the implied conditional variance of the CDX term premium. Since this conditional variance depends on the two Markov chain processes, combining the filtered probabilities of states $\sum_{S_{1,t-1}=0}^1 \sum_{S_{1,t}=0}^1 \sum_{S_{2,t-1}=0}^1 \sum_{S_{2,t}=0}^1 \Pr(S_{1,t-1}=i, S_{1,t}=j, S_{2,t-1}=i, S_{2,t}=j | I_{t-1})$ with

Equation (VII.21), we can calculate the conditional variance as a product of these two terms, based on the available information at time $t-1$. From Figure VII-6 we observe that the conditional variance, in the immediate aftermath of Bear Stern's bailout (March 14, 2008), remains at low levels for a few days. What is particularly striking is that the filtered probability of the high volatility regime for the RW component falls back to a near-zero value at the outbreak. In the sub-sample period surrounding Lehman Brothers' default, the conditional variance of the CDX term premium is much bigger than in the Bear Stern's bailout period, as a consequence, the state probability of the high volatility regime for RW component initially falls back to near-zero value, but rebounds very rapidly in the subsequent days as investors begin to worry about the stability of other systemically important financial institutions. Our results show that the coherence between the high conditional variance of the CDX term premium and the state probability of the high volatility regime for RW component disappears during the financial crisis period. This suggests that the RW and the stationary components may behave differently depending on whether the financial system is experiencing a systemic crisis or not.

VII.4.3 VAR Analysis

We now test for the impact of observed economic and financial variables on the unobserved stationary and RW components, in the context of the following VAR (4) model⁴⁸

⁴⁸ The order of the VAR is established using the SBC criterion.

$$Y_t = c + \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \Pi_3 Y_{t-3} + \Pi_4 Y_{t-4} + \varepsilon_t \quad (\text{VII.22})$$

where Y_t includes the stationary (*STAT*) and changes in the RW components (*DIFRW*), the observed level of effective US Federal Fund Rate (*FFR*), the slope of the yield curve (*SLOPE101*, calculated as the difference between 10-year and 1-year US treasury bond yields), the Standard & Poor's 500 index return (*SP500RTN*), and the implied volatility of the Standard & Poor's 500 index (*LOGVIX*)⁴⁹. c is a (6×1) vector of constants, Π_i are (6×6) coefficient matrices and ε_t is an (6×1) unobservable zero mean white noise vector process with invariant covariance matrix Σ . Given a structural break happened in January 2008, we split the sample into two periods depending on the two volatility regimes of the stationary component. The first sub-sample varies from September 13th, 2004 to January 3rd, 2008 and the second sub-sample starts from January 4th, 2008 to July 23rd, 2009.

To evaluate the relationships between the two components and monetary policy and stock market variables, we compute the accumulated generalized impulse response functions, which trace out the responsiveness of the dependent variables to one unit generalized shock to each of the variables in the VAR system. These impulse response functions provide useful insights into the dynamic properties of the system.

As forecasting variables of the variation in the term premium of the CDX index, these monetary policy and stock market variables appear to impact differently on the two components, before and after the onset of the crisis. Figure VII-7 shows the accumulated generalized impulse response functions of the VAR model in the pre-crisis period. The responses of *DIFRW* and *STAT* to *FFR* are significantly positive, which is consistent with the findings of Longstaff and Schwartz (1995) who document that corporate yield spreads vary inversely with the benchmark short-term treasury yield. Since an increase in the monetary policy rate would translate into an increase in the level of interest rate in normal periods, a higher interest rate level will decrease the present value of future cash flows and hence the value of default protection. This would lead to a tightening of the credit spread. If the tightening of the spread is more severe for the shorter CDS maturities, the term premium in the credit curve would increase as represented by a rise in the slope of the credit curve. This result, as illustrated in Figure VII-7, is consistent with Longstaff and Schwartz (1995)

⁴⁹ $Y_t = [FFR_t, SLOPE101_t, SP500RTN_t, LOGVIX_t, DIFRW_t, STAT_t]'$

who suggests that the relationship between the CDX index spread and the risk-free interest rate depends on the time horizon. Compared to the negative responses of DIFRW and STAT in Figure VII-8, an increase in the monetary policy rate in the crisis-period would sharply reduce liquidity, increasing the probability of imminent default. The obvious effect would be widening the 5-year CDX spread and flattening the term premium of the CDX index.

The responses of the two components to SLOPE101 are negative but insignificant during the pre-crisis period. Bedendo, Cathcart and El-Jahel (2007) report a negative relationship between the slope of the treasury yield curve and credit spreads. The reason is that when a positively sloped yield curve is the outcome of an expansionary monetary policy, which will increase future firm value and reduce default risk, then the term premium decreases due to falls in the 10-year spread. At the same time a steeper yield curve in normal time periods may indicate an increase in the future short rate through the expectations channel. If the inflation risk premium on longer term interest is very low (or even negative, see (Kim and Wright (2005b))), long term interest rates would change little in response to a continuous rise in short term rates (Smith and Taylor (2009)). This effect, known as the “Conundrum”, may make the shorter maturity default protection contract more expensive than those with longer maturities, and hence, increases the 5-year credit risk premium, reducing the term premium of CDX index. In contrast during the crisis time period, any increase in the slope of yield curve that caused by reduction in short rate would be regarded as a signal of liquidity providing from central banks in order to ameliorate the impact of the recession on the balance sheets of the firms. Default protection premia (5-year) will fall and this would subsequently lead to an increase in the term premium. The positive responses of both DIFRW and STAT to an increase in the slope of yield curve in crisis period (Figure VII-8) lend support to this hypothesis.

In the pre-crisis period the initial responses to equity market returns and the implied volatility index, have the opposite signs on DIFRW and STAT at the outset. Yet, at longer lags, both components exhibit a positive (negative) relationship with the equity returns (equity volatility). During the post-crisis period this initial divergence dissipates and both factors strongly respond positively to equity returns and negatively to the VIX. Such finding is consistent with the evidence in Byström (2006) who reports that on-the-run single-name CDS spread is significantly negatively related to the equity returns for the period 2004-2006. Scheicher (2009) also demonstrates that there is a significant contemporaneous link between the CDS market and the stock market. The inverse relationship between the two components

and equity return volatility is broadly consistent with previous econometric evidence, as illustrated by Campbell and Taksler (2003), Alexander and Kaeck (2008) and Zhang (2008). In the theoretical framework of Merton (1974), higher equity volatility means higher probability of hitting the default barrier, which induces a higher compensation on holding the bond in the form of larger credit spread. Although the positive (negative) responses to equity return (equity return volatility) hold both in non-crisis and crisis periods, the magnitude of responses in crisis period is far more pronounced.

To quantify the impact of all the observed variables on the unobserved factor we compute the generalized variance decomposition⁵⁰ over the two periods and the results are presented on Table VII-8 and Table VII-9. The key finding is the presence of strong effects attributable to stock market variables on both components before and after the crisis. Initially, in the pre-crisis period the impact of all the variables on both components is modest (Table VII-8). Stock market returns and their volatility appear to be the variables exerting some influence on DIFRW and STAT. Collectively they account for 7.5% of DIFRW variability's and 3.51% of STAT's variance.

Table VII-8: Generalized variance decomposition of DIFRW and STAT in the pre-financial crisis period(September 13, 2004 to January 03, 2008)

Variance Decomposition of DIFRW:							
Period	S.E.	FFR	SLOPE101	SP500RTN	LOGVIX	DIFRW	STAT
10	0.838	0.695	2.746	0.446	3.284	92.716	0.113
15	0.839	0.698	2.758	0.505	3.307	92.615	0.116
Variance Decomposition of STAT:							
10	0.065	0.506	0.579	1.209	0.875	4.320	92.512
15	0.065	0.509	0.579	1.445	0.973	4.307	92.186

In the post-crisis period of our sample (Table VII-9), approximately 13.5% of the variation in DIFRW can attributed to a combination of stock market variables, the monetary policy and the slope of the yield curve. More specifically the proportion accounted by the stock market variables is 7.68%. For the stationary component the same information set explains 16.44% of the variance a nearly fivefold increase compared to the previous period. The impact of monetary policy alone is 4.2% compared to its previous value of 0.5%.

⁵⁰ See Koop, Pesaran and Potter (1996) and Pesaran and Shin (1998)

Importantly, our results indicate that in the crisis period information pertinent to the immediate valuation of firm's assets such as the stock market index and monetary policy exerted relatively strong influence on the term premium, compared to factors such as VIX and the slope of the yield curve.

Within the confines of the model, the crisis period is indicative of a significant change to the fundamentals that determine the term premium, as captured by the stationary component. The evidence from the RW component is indicative of increased volatility that can be only partly explained by the set of the observed macro and financial variables included in our analysis. In summary, these results do provide evidence that the inversion of the term premium since the onset of the crisis is primarily attributable to the evolution of the stock market and monetary policy. In particular, the drastic and immediate response of the Fed played an important role in avoiding dramatic increases in the short (5-year) premia and therefore helped stabilizing the short-term segment of the CDS market.

Table VII-9: Generalized variance decomposition of DIFRW and STAT in the post-financial crisis period (January 04, 2008 to July 23, 2009)

Variance Decomposition of DIFRW:							
Period	S.E.	FFR	SLOPE101	SP500RTN	LOGVIX	DIFRW	STAT
⋮							
10	1.249	1.284	4.203	6.314	1.090	86.421	0.688
15	1.254	1.577	4.174	6.556	1.127	85.794	0.773
Variance Decomposition of STAT:							
⋮							
10	1.381	2.837	1.213	9.592	0.210	8.664	77.485
15	1.538	4.258	1.171	10.837	0.173	7.502	76.059

VII.5 Conclusion

In this article, we estimate a Markov switching unobserved component model to explain the evolution of the term premium of the most liquid CDS maturities for the North American CDX index.

We consider an appropriately specified Markov Switching Unobserved Components model as a reliable measure of volatility dynamics of the CDX index spread curve and investigate the presence and significance of both monetary policy adjustments and stock market returns for the US economy over the sample period September 2004 - July 2009.

To the best of our knowledge, this is the first direct empirically based evidence that is brought on the evolution of the term premium of the CDS index market and its observed macroeconomic and financial determinants.

To capture the magnitude of uncertainty in the credit risk transfer market, we decompose the level of the CDX index term premium into two components. The first, the RW component is assumed to capture the changes in volatility driving the term premium whereas the second, a stationary AR(1) process, represents the fundamentals. Furthermore, we formulate a model with time-varying regime switching probabilities and regime dependent components.

Our results suggest that the inversion of the curve around September 2008 is largely driven by abrupt moves in the stationary component, representing the evolution of the fundamentals under-pinning the probability of default in the economy. The component enters the high volatility regime after a prolonged period of remarkable stability. Notably, by the end of 2007, the stationary component exhibits slight turbulence in the low volatility regime, but of a very different order of magnitude from the subsequent evolution of the component in 2008. The decline of the term premium accelerates sharply throughout the end of 2007 when it partially reverses its trend remains. However, it remains in the high volatility regime in the final part of the sample period. Interestingly, although the RW component appears to evolve in a very stable and predictable manner from 2004 to 2008, fluctuates somewhat intensively between the low and high volatility regime over short periods of time during 2005 and by the end of 2007. From the beginning of the sub-prime crisis in August 2007 the component exhibits downward movement but does not enter decisively the high volatility regime. Over the last part of the sample, the component enters more frequently the low-volatility regime but in a rather unpredictable manner, indicating that the uncertainty surrounding asset values still remain unabated.

Remarkably, the inclusion of observed economic and financial variables to predict the evolution of the unobserved components does a relatively good job only during the ‘crisis’ period. These variables are found to make a statistically significant contribution that is consistent with economic theory. Indeed, we find robust evidence that the unprecedented monetary policy response, of sharp rate reductions by the Fed during the crisis period, was effective in reducing market uncertainty and helped to steepen the curve of the index thereby mitigating systemic risk concerns. The impact of stock market volatility in flattening the

curve and exerting comparatively higher upward pressure on the 5-year CDX is substantially more robust in the crisis period, as both components are significantly affected by the VIX measure. It also appears that equity returns are important drivers of the term premium during both periods. This impact results in a steepening of the curve as the current value of the underlying ‘collateral’ increases. Our results also suggest that in the pre-crisis period the RW component associated with increased volatility displays a low reaction to the stock market. Additionally, as expected the impact on the stationary component albeit positive is not significant. Yet, from January 01, 2008 both components respond immediately and significantly to stock market fluctuations. We demonstrate that the unprecedented stock market collapse is a very important contributory factor to the inversion of the CDX index term premium.

Overall, this evidence implies that credit risk modeling that ignores this regime dependent feature would bias the pricing of credit contracts. Developments in both the first and second moments of the equity market have a lasting influence on both components, with more pronounced effects during volatile market conditions.

The evolution of the CDX index in all maturities is an important signal of the ‘health’ of the economy over the short and long run. Sudden inversions indicate sharp deterioration of the current economic conditions and increased probability of default. Such movements are triggered by both the evolving stance of monetary policy and developments in the equity markets that make a significant albeit modest contribution to their predictability.

This article is only a first step toward the development of a fully fledged consistent framework to gain greater insight in the dynamics of the CDX curve indices across different parts of the credit cycle and in the relationship between the shape of the term structure and macro/financial variables fluctuations. Further research is warranted. Interesting possibilities for further research include the consideration of an extended number of maturities and of other index tranches, such the high-yield segment of the market. These extensions along with a complementing examination of liquidity risks and the risk of spillovers will enhance our understanding of the dynamics of such important markets, primarily from a systemic viewpoint.

Bibliography

- Alexander, Carol, and Andreas Kaeck, 2008, Regime dependent determinants of credit default swap spreads, *Journal of Banking & Finance* 32, 1008-1021.
- Ang, Andrew, and Monika Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* 50, 745-787.
- Aunon-Nerin, D, D Cossin, T Hricko, and Z. Huang, 2002, Exploring for the determinants of credit risk in credit default swap transaction data: Is fixed-income markets information sufficient to evaluate credit risk?, *HEC Working Paper*.
- Barclays Capital Research, Quantitative Credit Strategy, 2008b. *Cds curve trading handbook 2008*.
- Bedendo, Mascia, Lara Cathcart, and Lina El-Jahel, 2007, The slope of the term structure of credit spreads: An empirical investigation, *Journal of Financial Research* 30, 237-257.
- Bernanke, Ben, and Mark Gertler, 1989, Agency costs, net worth, and business fluctuations, *American Economic Review* 79, 14.
- Bernanke, Ben, Mark Gertler, and Simon Gilchrist, 1996, The financial accelerator and the flight, *Review of Economics & Statistics* 78, 1.
- Bernanke, Ben S., 1981, Bankruptcy, liquidity, and recession, *American Economic Review* 71, 155.
- Bernanke, Ben S., 1983, Nonmonetary effects of the financial crisis in the propagation of the great depression, *American Economic Review* 73, 257.
- Bhar, R., D. Colwell, and P. Wang, 2008, Component structure of credit default swap spreads and their determinants, Australasian Finance and Banking Conference 2008.
- Blanco, Roberto, Brennan Simon, and Ian W. Marsh, 2005, An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps, *The Journal of Finance* 60, 2255-2281.
- Byström, Hans, 2006, Creditgrades and the itraxx cds index market, *Financial Analysts Journal* 62, 65-76.
- Byström, Hans N. E., 2005, Using credit derivatives to compute marketwide default probability term structures, *Journal of Fixed Income* 15, 34-41.
- Campbell, John Y., and Glen B. Taksler, 2003, Equity volatility and corporate bond yields, *Journal of Finance* 58, 2321-2349.
- Clark, Peter K., 1987, The cyclical component of u. S. Economic activity, *The Quarterly Journal of Economics* 102, 797-814.
- Engle, Robert, and Mark Watson, 1981, A one-factor multivariate time series model of metropolitan wage rates, *Journal of the American Statistical Association* 76, 774-781.
- Fama, Eugene F., and Kenneth R. French, 1988, Permanent and temporary components of stock prices, *The Journal of Political Economy* 96, 246-273.
- Garcia, R., 1998, Asymptotic null distribution of the likelihood ratio test in markov switching models, *International Economic Review* 39, 763-788.
- Hamilton, James D, 1994b, State space models, in Rober Engel, and Dan McFadden, eds.: *Handbook of econometrics* (Elsevier).
- Hamilton, James D., 1989a, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357-384.
- Hansen, B. E., 1992, The likelihood ratio test under nonstandard conditions: Testing the markov switching model of gnp, *Journal of Applied Econometrics* 7, S61-S82.
- Harvey, A. C., 1985, Trends and cycles in macroeconomic time series, *Journal of Business & Economic Statistics* 3, 216-227.

- Harvey, A. C., 1989a. *Forecasting structural time series models and the kalman filter* (Cambridge University Press, Cambridge).
- Harvey, A.C., and N. Shephard, 1993, Structural time series models, in et al. G.S. Maddala, ed.: *Handbook of statistics* (Elsevier Science Publisher).
- Hull, John, Mirela Predescu, and Alan White, 2004, The relationship between credit default swap spreads, bond yields, and credit rating announcements, *Journal of Banking & Finance* 28, 2789-2811.
- Kim, Chang-Jin, 1994, Dynamic linear models with markov-switching, *Journal of Econometrics* 60, 1-22.
- Kim, Chang-Jin, and Charles R. Nelson, 1989, The time-varying-parameter model for modeling changing conditional variance: The case of the lucas hypothesis, *Journal of Business & Economic Statistics* 7, 433-440.
- Kim, Chang-Jin, Jeremy M. Piger, and Richard Startz, 2007, The dynamic relationship between permanent and transitory components of u.S. Business cycles.
- Kim, Don H., and Jonathan H. Wright, 2005b, An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates, (Board of Governors of the Federal Reserve System (U.S.)).
- Kim, Myung Jig, Charles R. Nelson, and Richard Startz, 1991, Mean reversion in stock prices? A reappraisal of the empirical evidence, *The Review of Economic Studies* 58, 515-528.
- Koop, G., M. H. Pesaran, and S. M. Potter, 1996, Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* 74, 119-147.
- Lamoureux, Christopher G., and William D. Lastrapes, 1990, Persistence in variance, structural change, and the garch model, *Journal of Business & Economic Statistics* 8, 225-234.
- Lo, A. W., and A. C. MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies* 1, 41-66.
- Lo, Andrew W., and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431-467.
- Longstaff, Francis A., Sanjay Mithal, and Eric Neis, 2005, Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market, *The Journal of Finance* 60, 2213-2253.
- Longstaff, Francis A., and Eduardo S. Schwartz, 1995, A simple approach to valuing risky fixed and floating rate debt, *Journal of Finance* 50, 789-819.
- Merton, Robert C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* 29, 449-470.
- Morley, James, and Jeremy Piger, 2008, Trend/cycle decomposition of regime-switching processes, *Journal of Econometrics* 146, 220-226.
- Nervole, M., D.M. Grether, and J.L. Carvalho, 1979. *Analysis of economic time series: A synthesis* (Academic Press, New York).
- Pan, Jun, and Kenneth J. Singleton, 2008, Default and recovery implicit in the term structure of sovereign cds spreads, *The Journal of Finance* 63, 2345-2384.
- Pesaran, H. H., and Y. Shin, 1998, Generalized impulse response analysis in linear multivariate models, *Economics Letters* 58, 17-29.
- Poterba, J. M., and L. H. Summers, 1988, Mean reversion in stock prices - evidence and implications, *Journal of Financial Economics* 22, 27-59.
- Scheicher, M., 2009, The correlation of a firm's credit spread with its stock price: Evidence from credit default swaps, in G. N. Gregoriou, ed.: *Stock market volatility* (Chapman & Hall/CRC Finance Series).
- Smith, J. M., and J. B. Taylor, 2009, The term structure of policy rules, *Journal of Monetary Economics* 56, 907-917.

- Stulz, Rene M., 2009, Credit default swaps and the credit crisis, (SSRN).
- Summers, Lawrence H., 1986, Does the stock market rationally reflect fundamental values?, *Journal of Finance* 41, 591-601.
- Taylor, John B., and John C. Williams, 2009, A black swan in the money market.
- Watson, Mark W., 1986, Univariate detrending methods with stochastic trends, *Journal of Monetary Economics* 18, 49-75.
- Zhang, Frank Xiaoling, 2008, Market expectations and default risk premium in credit default swap prices: A study of argentine default, *Journal of Fixed Income* 18, 37-55.
- Zhu, Haibin, 2006, An empirical comparison of credit spreads between the bond market and the credit default swap market, *Journal of Financial Services Research* 29, 211-235.

Appendix VII

Figure VII-4: Stationary component and filtered state probability of its high volatility regime

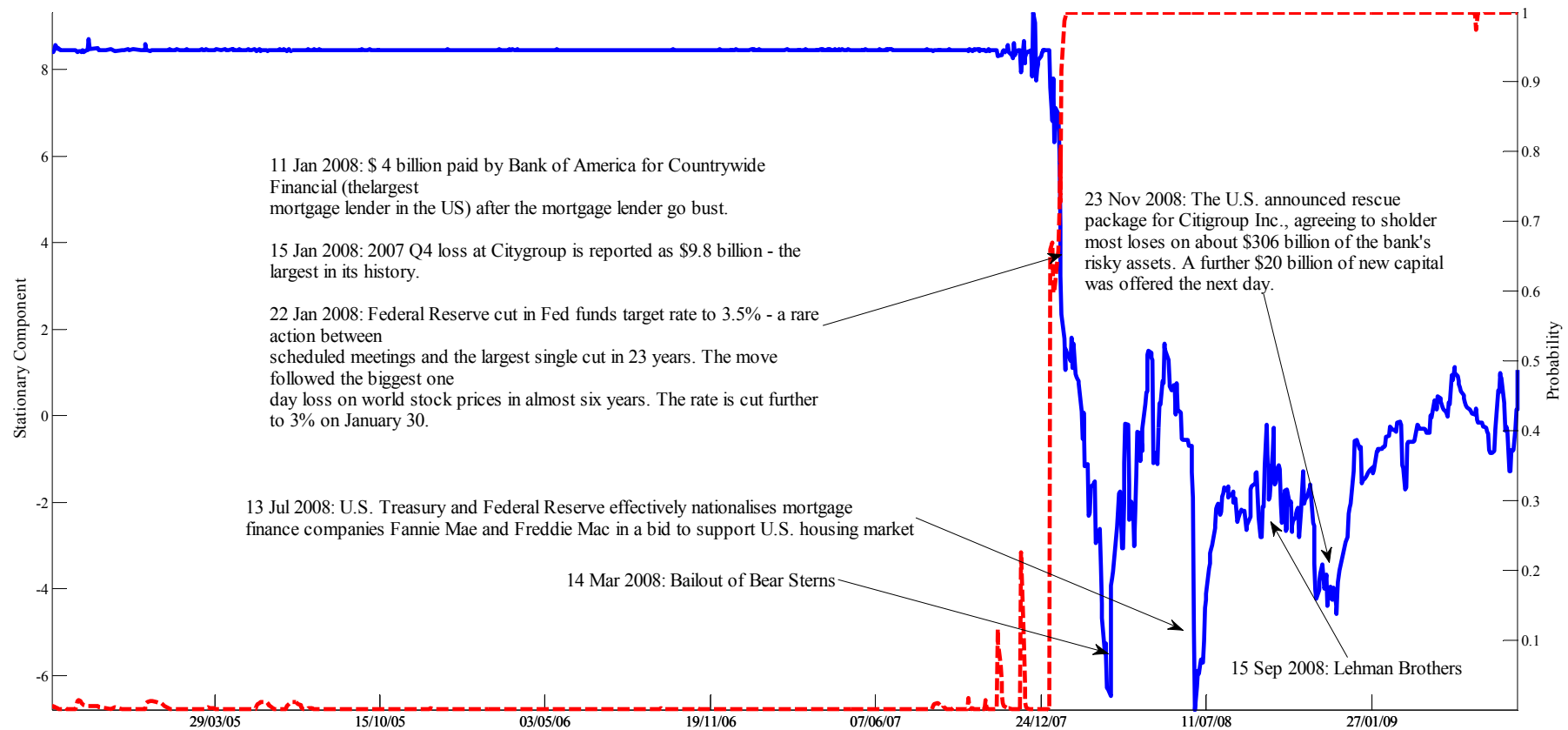


Figure VII-5: Random walk component and filtered state probability of its high volatility regime

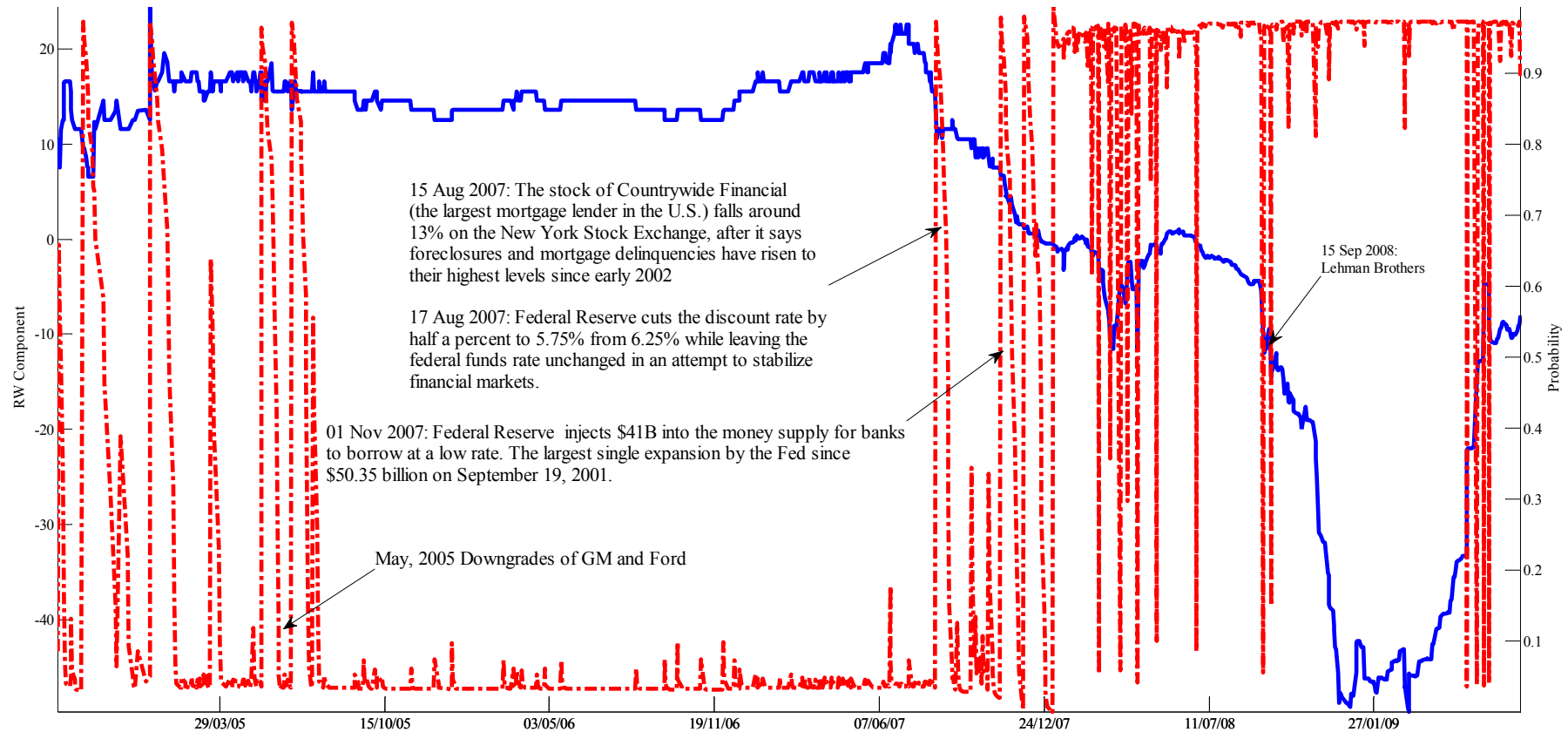


Figure VII-6: Conditional variance of the term premium and filtered state probability of high volatility regime for each component

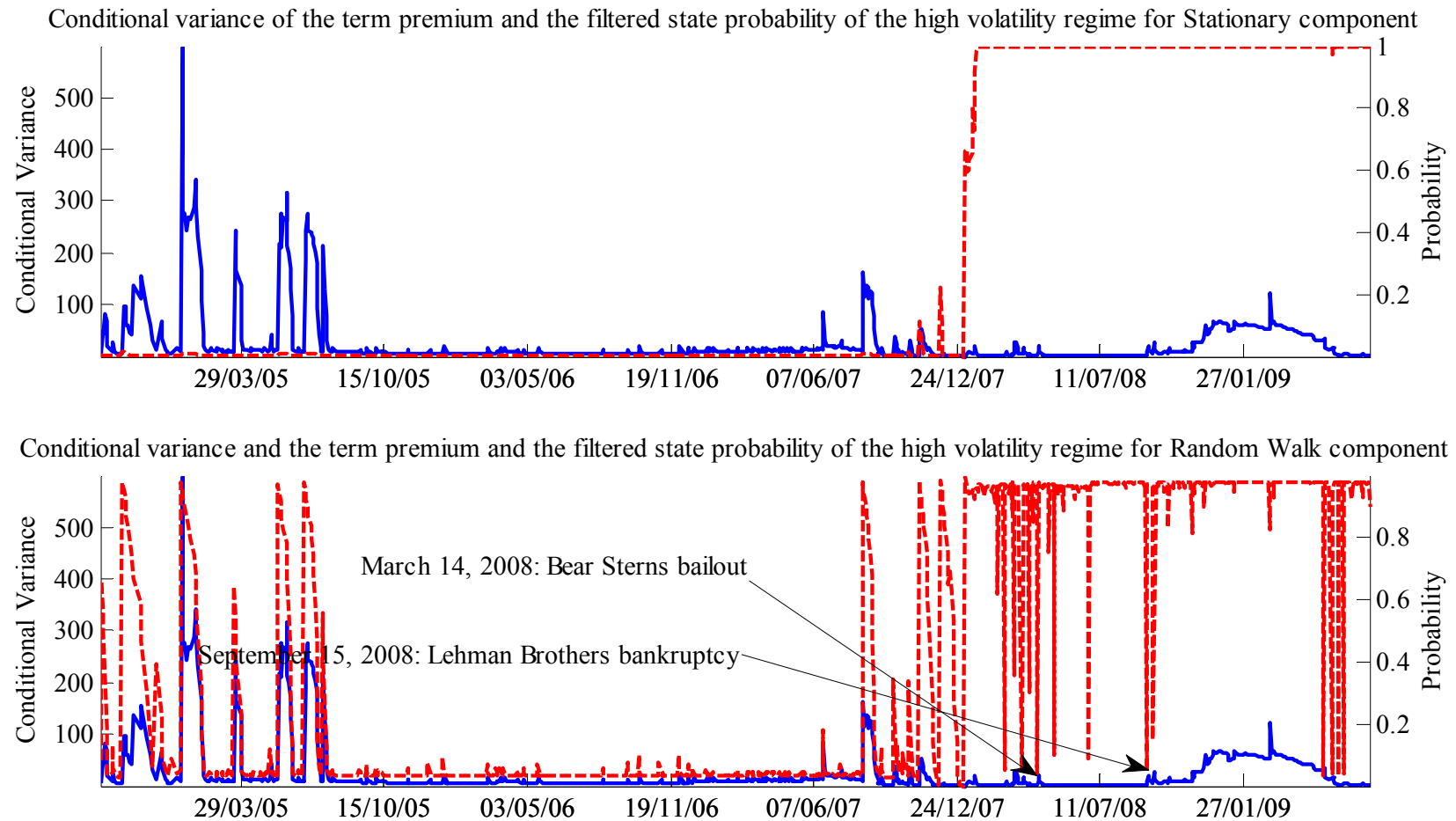


Figure VII-7: Accumulative generalized impulse response functions of the VAR model in the pre-financial crisis period (September 13, 2004 to January 03, 2008)

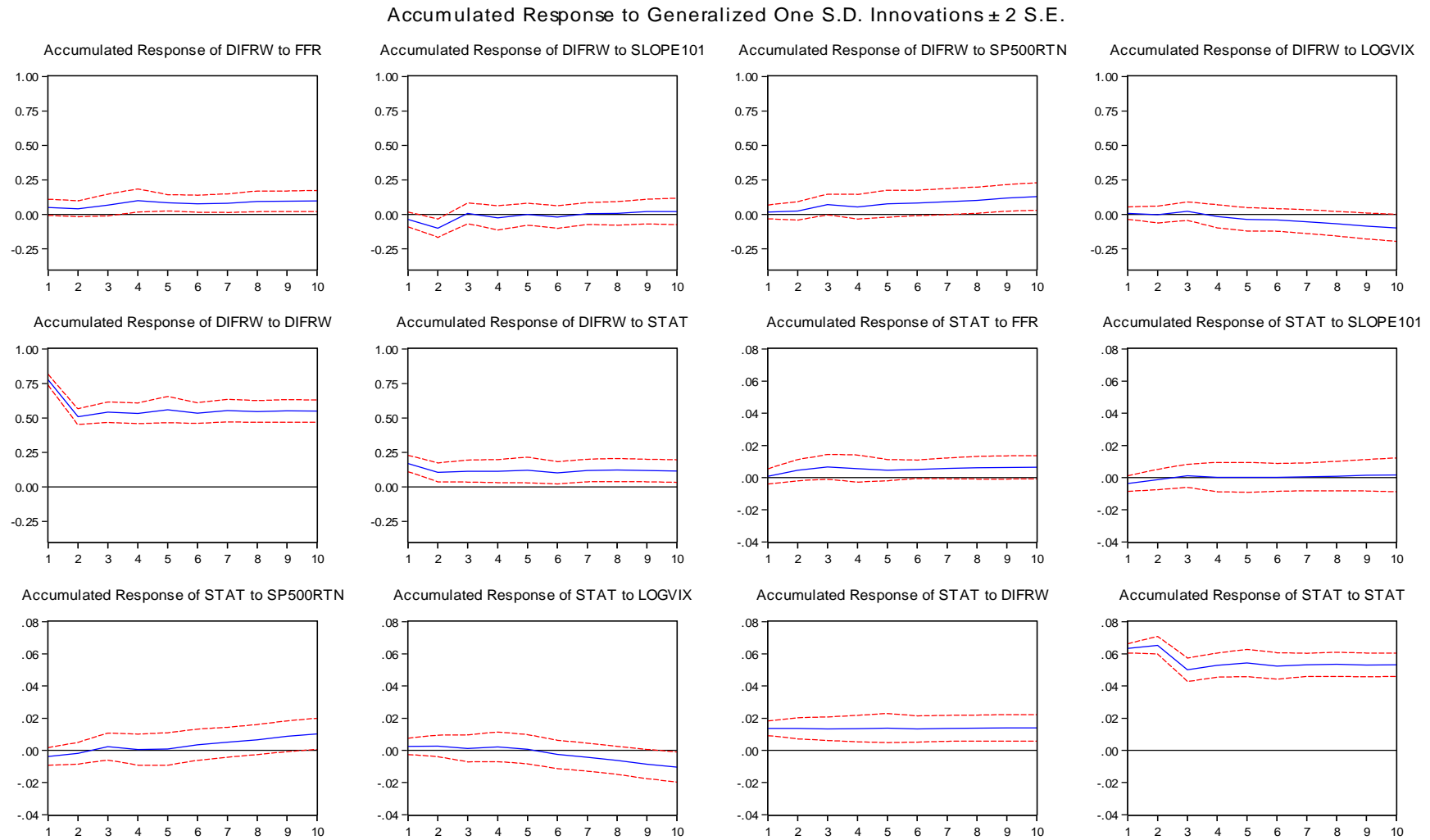
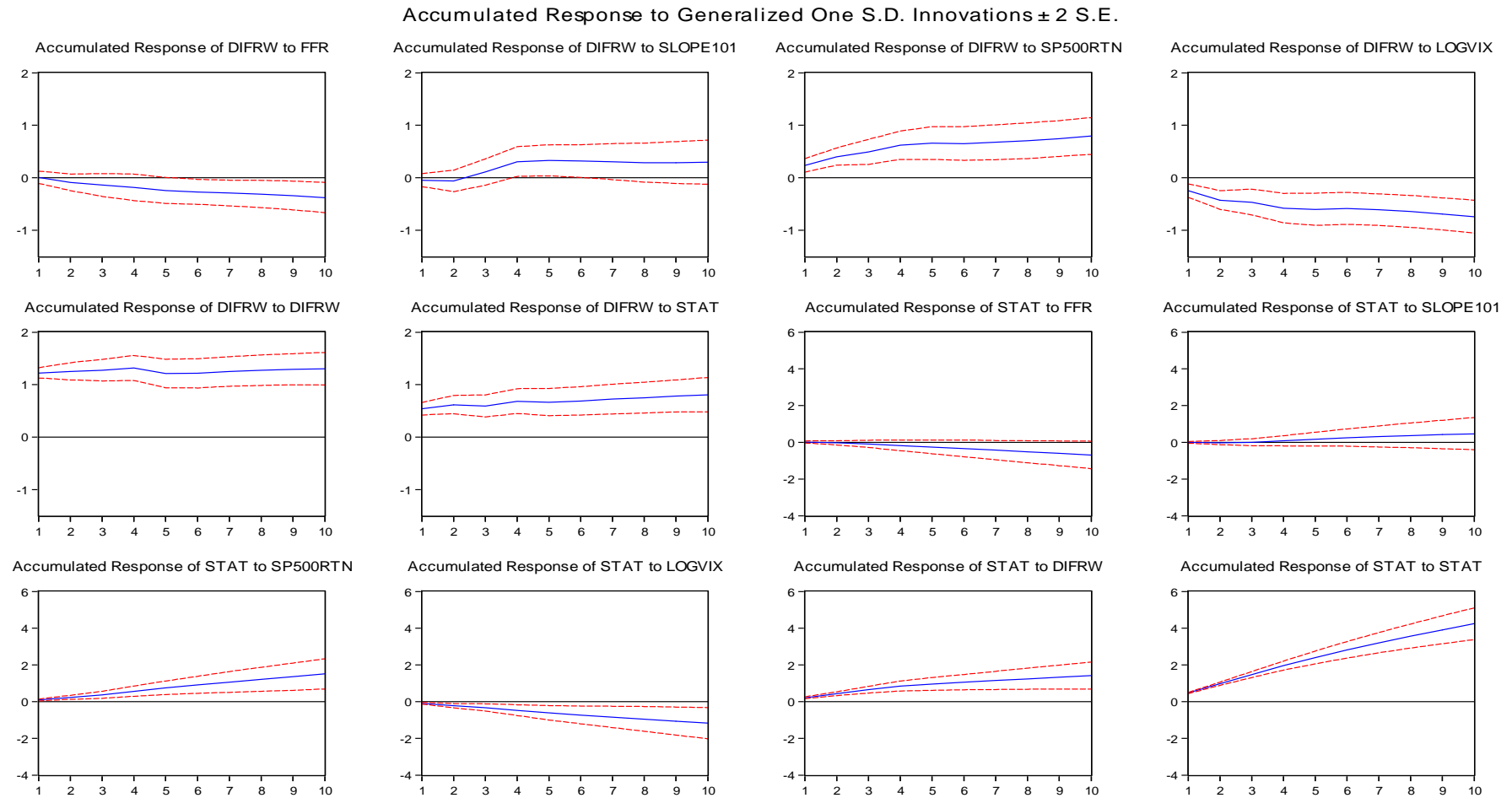


Figure VII-8: Accumulative generalized impulse response functions of the VAR model in the post-financial crisis period (January 04, 2008 to July 23, 2009)



Matlab Code (Model 8)

The following codes require CompEcon Toolbox to run. The toolbox can be downloaded from www4.ncsu.edu/~pfackler/compecon

```
%% =====  
% This is the main file to run the Markov switching unobserved components  
% model  
% =====  
clear all; close all;  
global yy t ffr_ch slope101 rtn spvol  
%%% DATA  
load longdataTV2.txt;  
data=longdataTV2;  
date_t=data(:,1);  
tp=data(:,2);  
ffr_ch=data(:,3);  
slope101=data(:,7);  
rtn=data(:,4);  
spvol=data(:,6);  
% construct variables  
date_t=x2mdate(date_t);  
yy=tp;  
t=size(yy,1);  
  
%% ESTIMATION  
% parameters input  
para_in=[8.00519897691168;20.6248206446350;  
-21.4452891230563;1.38555671854035;-3.87375316140109;1.86923218059667;  
0.113664191935447;-0.106652956634136;-0.0347963549804958;-0.0131276023921204;  
3.37045547055373;3.26087476061801;5.48696462364791;13.2059955018633;  
4.77946765585880;-0.544974875957186];  
%% optimization  
options=optimset('Display','iter','TolX',1e-4,'TolFun',1e-8,...  
'MaxFunEvals',1000000,'MaxIter',100000);  
  
[xout,fval,exitflag,output]=fminsearch(@likhfcn,para_in,options);  
  
%% RESULTS  
% parameters  
xfnl=trans(xout);  
% S.E. and t-test  
h_0=fdhess(@likhfcn,xout); % calculate Hessian matrix  
g_0=fdjac1(@trans,xout,[]);  
h_fnl=g_0*(inv(h_0))*g_0'; % final Hessian  
std_fnl=sqrt(diag(h_fnl)); % S.E.  
t_ratio=xfnl./std_fnl;  
  
% print results  
sprintf('=====')  
sprintf('The Estimated Parameters are:')  
disp(xfnl)  
sprintf('Standardized Errors of the Parameters are:')  
disp(std_fnl)  
sprintf('=====')  
sprintf('t-ratio for Estimated Parameters are:')
```



```

disp(t_ratio)
% Kalman and Smooth FILTER
% Kalman
[X_hat,Pw_hat,Pv_hat,pr_TL_w_1,pr_TL_v_1,sd_resid,var_pred_e,pred_e]=kfilter(xout);
% plot the two factors and probabilities
figure(1)
plot(X_hat);
figure(2)
plot(Pw_hat,'b');
figure(3)
plot(Pv_hat,'r');
figure(4)
plot(sd_resid(20:end),'k');
figure(5)
plot(var_pred_e(20:end),'c');
figure(6)
plot(pred_e(20:end),'g');

% smooth
[sPw_1,sPv_1]=sfilter(xfml,Pw_hat,Pv_hat,pr_TL_w_1,pr_TL_v_1);
figure(7)
plot(sPw_1,'b');
figure(8)
plot(sPv_1,'r');

```

```

function likval=likhfcn(para_likhfcn)

```

```

%=====
% Likelihood function returns likval to be optimized
%=====

```

```

global yy t ffr_ch slope101 rtn spvol

```

```

%% assign parameters

```

```

para=trans(para_likhfcn);
delta_t=1/250;
delta=para(1);
k2=para(15);
k=para(2);
delta2=para(16);
% valatility
w_0=para(3)*sqrt(delta_t);
w_1=para(4)*sqrt(delta_t);
v_0=para(5)*sqrt((1-exp(-2*k*delta_t))/(2*k));
v_1=para(6)*sqrt((1-exp(-2*k2*delta_t))/(2*k2));

```

```

% correlation

```

```

corr_w0v0=para(7);
corr_w0v1=para(8);
corr_w1v0=para(9);
corr_w1v1=para(10);

```

```

% initial transition probability

```

```

p00=para(11); % Pr[Sw(t)=0 | Sw(t-1)=0]
p11=para(12); % Pr[Sw(t)=1 | Sw(t-1)=1]
q00=para(13); % Pr[Sv(t)=0 | Sv(t-1)=0]
q11=para(14); % Pr[Sv(t)=1 | Sv(t-1)=1]

```

```

%% construct matrices

```

```

H=[1 1]; % measurement

```

```

C1=[0; delta*(1-exp(-k*delta_t))];
C2=[0; delta2*(1-exp(-k2*delta_t))];
F1=[1 0;0 exp(-k*delta_t)]; % transition matrix
F2=[1 0;0 exp(-k2*delta_t)];
% covariance matrices
cov_w0v0=[w_0^2      corr_w0v0*w_0*v_0;
          corr_w0v0*w_0*v_0  v_0^2];

cov_w0v1=[w_0^2      corr_w0v1*w_0*v_1;
          corr_w0v1*w_0*v_1  v_1^2];

cov_w1v0=[w_1^2      corr_w1v0*w_1*v_0;
          corr_w1v0*w_1*v_0  v_0^2];

cov_w1v1=[w_1^2      corr_w1v1*w_1*v_1;
          corr_w1v1*w_1*v_1  v_1^2];

Q_00=cov_w0v0;
Q_01=cov_w0v1;
Q_10=cov_w1v0;
Q_11=cov_w1v1;

%% steady state probabilities
ssp1=(1-p00)/(2-p00-p11); % Pr[Sw(t-1)=1|Y(t-1)]
ssp0=1-ssp1; % Pr[Sw(t-1)=0|Y(t-1)]
ssq1=(1-q00)/(2-q00-q11); % Pr[Sv(t-1)=1|Y(t-1)]
ssq0=1-ssq1; % Pr[Sv(t-1)=0|Y(t-1)]

%% initial values for factors and their variances
X_00=zeros(2,1); % w0v0
X_01=zeros(2,1); % w0v1
X_10=zeros(2,1); % w1v0
X_11=zeros(2,1); % w1v1

prior=[100000 0;0 100000];
P_00=prior; %w0v0
P_01=prior; %w0v1
P_10=prior; %w1v0
P_11=prior; %w1v1

%% create space to store values
X_fore = zeros(2,16);
P_fore = zeros(2,2*16);
yy_error = zeros(1,16);
Kalman_gain = zeros(2,16);
X_up = zeros(2,16);
P_up = zeros(2,2*16);
pr_vl = zeros(1,16);

%% create map and start the Kalman filter + Hamilton filter
likval=0;
for j_iter=1:t
    j=1;
    for Swt=0:1
        for Sv1=0:1
            for Swt_1=0:1
                for Sv1_1=0:1
                    if Swt_1==0 && Sv1_1==0

```

```

X_=X_00;
P_=P_00;

elseif Swt_1==0 && Svt_1==1
    X_=X_01;
    P_=P_01;
elseif Swt_1==1 && Svt_1==0
    X_=X_10;
    P_=P_10;
elseif Swt_1==1 && Svt_1==1
    X_=X_11;
    P_=P_11;
end
%-----
if Swt==0 && Svt==0
    C=C1;
    F=F1;
    Q_=Q_00;
elseif Swt==0 && Svt==1
    F=F2;
    Q_=Q_01;
    C=C2;
elseif Swt==1 && Svt==0
    C=C1;
    F=F1;
    Q_=Q_10;
elseif Swt==1 && Svt==1
    F=F2;
    Q_=Q_11;
    C=C2;
end
%-----
% start the Kalman filter
X_fore(:,j)=C+F*X_ ; % forecast for X
P_fore(:,(j-1)*2+1:(j-1)*2+2)=F*P_*F'+Q_ ; % forecast for variance
yy_error(j)=yy(j_iter)-H*X_fore(:,j); % forecast error
Kalman_gain(:,j)=P_fore(:,(j-1)*2+1:(j-1)*2+2)*H'*inv(H*P_fore(:,(j-1)*2+1:(j-1)*2+2)*H');
%-----
% updating
X_up(:,j)=X_fore(:,j)+Kalman_gain(:,j)*yy_error(j);
P_up(:,(j-1)*2+1:(j-1)*2+2)=P_fore(:,(j-1)*2+1:(j-1)*2+2)-Kalman_gain(:,j)*H'*P_fore(:,(j-
1)*2+1:(j-1)*2+2);
%-----
% now, include the transition probabilities
if Swt==1 && Swt_1==1
    trp = p11;
    trp_ = ssp1;
elseif Swt==1 && Swt_1==0
    trp = 1-p00;
    trp_ = ssp0;
elseif Swt==0 && Swt_1==1
    trp = 1-p11;
    trp_ = ssp1;
elseif Svt==0 && Svt_1==0
    trp = p00;
    trp_ = ssp0;
end
%-----
if Svt==1 && Svt_1==1

```

```

        trq = q11;
        trq_ = ssq1;
    elseif Svt==1 && Svt_1==0
        trq = 1-q00;
        trq_ = ssq0;
    elseif Svt==0 && Svt_1==1
        trq = 1-q11;
        trq_ = ssq1;
    elseif Svt==0 && Svt_1==0
        trq = q00;
        trq_ = ssq0;
    end
    %-----
    % calculate likelihood value
    pr_vl(j)=(1/sqrt(2*pi*det(H*P_fore(:,(j-1)*2+1:(j-1)*2+2)*H')))*...
        exp(-0.5*yy_error(j)*inv(H*P_fore(:,(j-1)*2+1:(j-1)*2+2)*H')*yy_error(j))*...
        trp*trp_*trq*trq_;
    j=j+1;
end
end
end
end
%-----
% inference on probability terms
pr_val=sum(pr_vl,2); % f(Y(t)|I(t-1))
lik=log(pr_val);
pro_=pr_vl/pr_val; % ratio of f(Y(t)|Swt,Swt_1,Svt,Svt_1)/f(Y(t)|I(t))
%-----
% calculate joint probabilities e.g. Pr[Sv1t=0,Sv1t_1=0|Y(t)]
% 4 by 1:
%first one: Pr[Swt=0,Swt_1=0|Y(t)]
%second one: Pr[Swt=0,Swt_1=1|Y(t)]
pro_w=ref_file1(pro_,1);
% 4 by 1:
%first one: Pr[Svt=0,Svt_1=0|Y(t)]
%second one: Pr[Svt=0,Svt_1=1|Y(t)]
pro_v=ref_file1(pro_,2);
% 4 by 1:
%first one: Pr[Swt=0,Svt=0|Y(t)]
%second one: Pr[Swt=1,Svt=0|Y(t)]
pro_wv=ref_file1(pro_,3);

ssp0=sum(pro_w(1:2));
ssp1=sum(pro_w(3:4));
ssq0=sum(pro_v(1:2));
ssq1=sum(pro_v(3:4));
%=====
%% Collapse Approximation
% factor collapse approximation
app_out=app_file1(pro_,X_up,1);
X_00=app_out(:,1)./pro_wv(1);
X_10=app_out(:,2)./pro_wv(2);
X_01=app_out(:,3)./pro_wv(3);
X_11=app_out(:,4)./pro_wv(4);

% factor variance collapse approximation
[fac_00, fac_10, fac_01, fac_11]=app_file2(pro_,P_up,app_out,X_up);
P_00(1,1)=fac_00(1,1)/pro_wv(1);
P_10(1,1)=fac_10(1,1)/pro_wv(2);

```

```

P_01(1,1)=fac_01(1,1)/pro_wv(3);
P_11(1,1)=fac_11(1,1)/pro_wv(4);

P_00(2,2)=fac_00(2,2)/pro_wv(1);
P_10(2,2)=fac_10(2,2)/pro_wv(2);
P_01(2,2)=fac_01(2,2)/pro_wv(3);
P_11(2,2)=fac_11(2,2)/pro_wv(4);

% update likelihood value
if j_iter>=20
    likval=likval-lik;
end
end
end % of function

function prob_out=ref_file1(prob_para,id_ms)
St_mat=zeros(4,16);
prob=prob_para;

j=1;
for Sv1t=0:1
    for Sv2t=0:1
        for Sv1t_1=0:1
            for Sv2t_1=0:1
                St_mat(:,j)=[Sv1t Sv2t Sv1t_1 Sv2t_1];
                j=j+1;
            end
        end
    end
end
end
% switch cases
ind_00=zeros(1,16);
ind_01=zeros(1,16);
ind_10=zeros(1,16);
ind_11=zeros(1,16);
switch id_ms
case 1
    for i=1:size(St_mat,2)
        if St_mat(3,i)==0 && St_mat(1,i)==0
            ind_00(i)=1;
        else
            ind_00(i)=0;
        end
        if St_mat(3,i)==1 && St_mat(1,i)==0
            ind_10(i)=1;
        else
            ind_10(i)=0;
        end
        if St_mat(3,i)==0 && St_mat(1,i)==1
            ind_01(i)=1;
        else
            ind_01(i)=0;
        end
        if St_mat(3,i)==1 && St_mat(1,i)==1
            ind_11(i)=1;
        else
            ind_11(i)=0;
        end
    end
end

```

```

end
prob_out=zeros(4,1);
prob_out(1)=prob*ind_00';
prob_out(2)=prob*ind_10';
prob_out(3)=prob*ind_01';
prob_out(4)=prob*ind_11';
%-----
case 2
for i=1:size(St_mat,2)
    if St_mat(4,i)==0 && St_mat(2,i)==0
        ind_00(i)=1;
    else
        ind_00(i)=0;
    end
    if St_mat(4,i)==1 && St_mat(2,i)==0
        ind_10(i)=1;
    else
        ind_10(i)=0;
    end
    if St_mat(4,i)==0 && St_mat(2,i)==1
        ind_01(i)=1;
    else
        ind_01(i)=0;
    end
    if St_mat(4,i)==1 && St_mat(2,i)==1
        ind_11(i)=1;
    else
        ind_11(i)=0;
    end
end
prob_out=zeros(4,1);
prob_out(1)=prob*ind_00';
prob_out(2)=prob*ind_10';
prob_out(3)=prob*ind_01';
prob_out(4)=prob*ind_11';
%-----
case 3
for i=1:size(St_mat,2)
    if St_mat(1,i)==0 && St_mat(2,i)==0
        ind_00(i)=1;
    else
        ind_00(i)=0;
    end
    if St_mat(1,i)==1 && St_mat(2,i)==0
        ind_10(i)=1;
    else
        ind_10(i)=0;
    end
    if St_mat(1,i)==0 && St_mat(2,i)==1
        ind_01(i)=1;
    else
        ind_01(i)=0;
    end
    if St_mat(1,i)==1 && St_mat(2,i)==1
        ind_11(i)=1;
    else
        ind_11(i)=0;
    end
end
prob_out=zeros(4,1);

```

```

    prob_out(1)=prob*ind_00';
    prob_out(2)=prob*ind_10';
    prob_out(3)=prob*ind_01';
    prob_out(4)=prob*ind_11';

    otherwise
        error('input id_ms must be 1 or 2 or 3');
    end

function fac_out=app_file1(prob_para,fac_inp,id_ms)
St_mat=zeros(4,16);
prob=prob_para;
dd=fac_inp;
j=1;
for Sv1t=0:1
    for Sv2t=0:1
        for Sv1t_1=0:1
            for Sv2t_1=0:1
                St_mat(:,j)=[Sv1t Sv2t Sv1t_1 Sv2t_1];
                j=j+1;
            end
        end
    end
end

ind_00=zeros(1,16);
ind_01=zeros(1,16);
ind_10=zeros(1,16);
ind_11=zeros(1,16);

switch id_ms
case 1
    for i=1:size(St_mat,2)
        if St_mat(2,i)==0 && St_mat(1,i)==0
            ind_00(i)=1;
        else
            ind_00(i)=0;
        end
        if St_mat(2,i)==0 && St_mat(1,i)==1
            ind_10(i)=1;
        else
            ind_10(i)=0;
        end
        if St_mat(2,i)==1 && St_mat(1,i)==0
            ind_01(i)=1;
        else
            ind_01(i)=0;
        end
        if St_mat(2,i)==1 && St_mat(1,i)==1
            ind_11(i)=1;
        else
            ind_11(i)=0;
        end
    end
    fac_out=zeros(2,4);
    fac_out(1,1)=(prob.*ind_00)*(dd(1,:).*ind_00)';

```

```

    fac_out(1,2)=(prob.*ind_10)*(dd(1,:).*ind_10)';
    fac_out(1,3)=(prob.*ind_01)*(dd(1,:).*ind_01)';
    fac_out(1,4)=(prob.*ind_11)*(dd(1,:).*ind_11)';
    %-----
    fac_out(2,1)=(prob.*ind_00)*(dd(2,:).*ind_00)';
    fac_out(2,2)=(prob.*ind_10)*(dd(2,:).*ind_10)';
    fac_out(2,3)=(prob.*ind_01)*(dd(2,:).*ind_01)';
    fac_out(2,4)=(prob.*ind_11)*(dd(2,:).*ind_11)';
otherwise
    error('input id_ms must be a positive number: 1');
end

function [fac_00,fac_10,fac_01,fac_11]=app_file2(prob_para,var_inp,pre_out,fac_inp)

prob=prob_para;
var=var_inp;
pre=pre_out;
dd=fac_inp;
St_mat=zeros(4,16);
j=1;
for Sv1t=0:1
    for Sv2t=0:1
        for Sv1t_1=0:1
            for Sv2t_1=0:1
                St_mat(:,j)=[Sv1t Sv2t Sv1t_1 Sv2t_1];
                j=j+1;
            end
        end
    end
end
end

%% the first factor
ind_00=zeros(1,4);
ind_01=zeros(1,4);
ind_10=zeros(1,4);
ind_11=zeros(1,4);

kk=1;
for i=1:size(St_mat,2)
    if St_mat(1,i)==0 && St_mat(2,i)==0
        ind_00(kk)=i;
        kk=kk+1;
    end
end
kk=1;
for i=1:size(St_mat,2)
    if St_mat(1,i)==0 && St_mat(2,i)==1
        ind_01(kk)=i;
        kk=kk+1;
    end
end
kk=1;
for i=1:size(St_mat,2)
    if St_mat(1,i)==1 && St_mat(2,i)==0
        ind_10(kk)=i;
        kk=kk+1;
    end
end
end

```



```

kk=1;
for i=1:size(St_mat,2)
    if St_mat(1,i)==1 && St_mat(2,i)==1
        ind_11(kk)=i;
        kk=kk+1;
    end
end

%%

fac_00=(prob(ind_00(1))*(var(:,(ind_00(1)-1)*2+1:(ind_00(1)-1)*2+2)+(pre(:,1)-dd(:,ind_00(1)))*(pre(:,1)-dd(:,ind_00(1)))))+...
    prob(ind_00(2))*(var(:,(ind_00(2)-1)*2+1:(ind_00(2)-1)*2+2)+(pre(:,1)-dd(:,ind_00(2)))*(pre(:,1)-dd(:,ind_00(2)))))+...
    prob(ind_00(3))*(var(:,(ind_00(3)-1)*2+1:(ind_00(3)-1)*2+2)+(pre(:,1)-dd(:,ind_00(3)))*(pre(:,1)-dd(:,ind_00(3)))))+...
    prob(ind_00(4))*(var(:,(ind_00(4)-1)*2+1:(ind_00(4)-1)*2+2)+(pre(:,1)-dd(:,ind_00(4)))*(pre(:,1)-dd(:,ind_00(4)))));

fac_01=(prob(ind_01(1))*(var(:,(ind_01(1)-1)*2+1:(ind_01(1)-1)*2+2)+(pre(:,2)-dd(:,ind_01(1)))*(pre(:,2)-dd(:,ind_01(1)))))+...
    prob(ind_01(2))*(var(:,(ind_01(2)-1)*2+1:(ind_01(2)-1)*2+2)+(pre(:,2)-dd(:,ind_01(2)))*(pre(:,2)-dd(:,ind_01(2)))))+...
    prob(ind_01(3))*(var(:,(ind_01(3)-1)*2+1:(ind_01(3)-1)*2+2)+(pre(:,2)-dd(:,ind_01(3)))*(pre(:,2)-dd(:,ind_01(3)))))+...
    prob(ind_01(4))*(var(:,(ind_01(4)-1)*2+1:(ind_01(4)-1)*2+2)+(pre(:,2)-dd(:,ind_01(4)))*(pre(:,2)-dd(:,ind_01(4)))));

fac_10=(prob(ind_10(1))*(var(:,(ind_10(1)-1)*2+1:(ind_10(1)-1)*2+2)+(pre(:,3)-dd(:,ind_10(1)))*(pre(:,3)-dd(:,ind_10(1)))))+...
    prob(ind_10(2))*(var(:,(ind_10(2)-1)*2+1:(ind_10(2)-1)*2+2)+(pre(:,3)-dd(:,ind_10(2)))*(pre(:,3)-dd(:,ind_10(2)))))+...
    prob(ind_10(3))*(var(:,(ind_10(3)-1)*2+1:(ind_10(3)-1)*2+2)+(pre(:,3)-dd(:,ind_10(3)))*(pre(:,3)-dd(:,ind_10(3)))))+...
    prob(ind_10(4))*(var(:,(ind_10(4)-1)*2+1:(ind_10(4)-1)*2+2)+(pre(:,3)-dd(:,ind_10(4)))*(pre(:,3)-dd(:,ind_10(4)))));

fac_11=(prob(ind_11(1))*(var(:,(ind_11(1)-1)*2+1:(ind_11(1)-1)*2+2)+(pre(:,4)-dd(:,ind_11(1)))*(pre(:,4)-dd(:,ind_11(1)))))+...
    prob(ind_11(2))*(var(:,(ind_11(2)-1)*2+1:(ind_11(2)-1)*2+2)+(pre(:,4)-dd(:,ind_11(2)))*(pre(:,4)-dd(:,ind_11(2)))))+...
    prob(ind_11(3))*(var(:,(ind_11(3)-1)*2+1:(ind_11(3)-1)*2+2)+(pre(:,4)-dd(:,ind_11(3)))*(pre(:,4)-dd(:,ind_11(3)))))+...
    prob(ind_11(4))*(var(:,(ind_11(4)-1)*2+1:(ind_11(4)-1)*2+2)+(pre(:,4)-dd(:,ind_11(4)))*(pre(:,4)-dd(:,ind_11(4)))));

```

Chapter VIII Summary and conclusion

This thesis studies the nonlinear relationships between financial (and economic) variables within the field of financial econometrics. It contains two reviews of literatures, one on nonlinear time series models and the other one on term structure of interest rates, and four empirical studies that explore various topics in fixed income, equity and credit markets.

In the first empirical essay, we provide comprehensive model specification tests under the frame of Markov switching CIR model on the term structure of UK interest rates. In contrast to Driffill, *et al.* (2009) which focuses on the US market, our study explores the UK market with a much longer sample period. We contribute to the literature by finding that the least restricted model provides the best in-sample estimation results for the term structure of UK interest rates. Although models with restrictive specifications may provide slightly better out-of-sample forecasts in directional movements of the yields, the economic gains seem to be small.

Unlike the first essay that looks into the nonlinear single factor model of the UK interest rates, we jointly model the nominal and real term structure of the UK interest rates using a three-factor essentially affine no-arbitrage term structure model in the second empirical study. In the first part of this essay, we decompose the nominal term structure of interest rates into real interest rates, expected inflation rates and inflation risk premia. Unlike previous literatures using observed variables in constructing expected inflation rates, we use latent factors to channel the dynamics of driving forces for the nominal pricing kernel while linking the real pricing kernel with the nominal one with an explicit modelling of the time-varying expected inflation rates. We contribute to the literature by finding that: (1) the smooth expected month-on-month inflation rate rather behaves like the trend extracted from the observed series; and on average the expected series is above the trend before the year of 2000, but remains below it in the subsequent periods; (2) the expected annual inflation rate also overstates the inflation rate before the new millennium and understates it in the following periods, which may suggest that the Bank of England's monetary policy is eventually quite creditable after it became independence in 1997; (3) the recent sharp decline of the expected inflation rates (which even lead to a deflation) may lend support to the standing ground of the central bank to keep interest rates at historically low level; (4) the 10-year inflation risk premia is relatively high since the inception of the 2007/2008 financial

crisis, and the possible explanation on this is that the inflation risk premia estimated for this particular sample period may contain a large amount of liquidity premia.

The second part of this essay investigates into the nonlinear relationship between monthly FTSE 100 index return risk premium and the expected inflation rates, which are filtered out previously. To our best knowledge, this study is the first one to address the relationship between stock return and inflation rates under a more advanced nonlinear modelling practice. The nonlinearity test based on a STVAR framework shows that there exists a nonlinear adjustment on the impact from lagged inflation rates to current return premium. The only one that has a significant forecasting effect on the return risk premium comes from the 4-period lagged changes of expected inflation rate. In one regime, one unit increase in the expected inflation rate forecasts a negative impact on the return premium; while in the other regime, a same unit increase in the expected inflation rate forecasts a positive adjustment on the return premium. The changes of regime, or the changes of sign on the impact of expected inflation rate to return premium is determined by the threshold value. While a larger change in the expected inflation rate forecasts a positive stock return premium, a modest change in the expected inflation rate predicts a negative response.

The third empirical study provides us additional insight into the nature of the aggregate stock market volatilities and its relationship to the expected returns, in a Markov switching model framework. Our work has benefited from using centuries-long aggregate stock market data from six countries (Australia, Canada, Sweden, Switzerland, UK and US). We contribute to the literature by finding that: (1) the Markov switching model assuming both regime dependent mean and volatility with a 3-regime specification is capable to captures the extreme movements of the stock market which are short-lived; (2) as most of the extreme movements in the stock market are downward drops, Markov switching models with a 3- regime specification would better capture the negative skewness in the stock returns than their 2-regime counterparts; (3) the volatility feedback effect that we studied on these six countries shows a positive sign on anticipating a high volatility regime of the current trading month; (4) having taken account for the volatility feedback effect, stock prices move in the opposite direction to the level of market volatility which creates an immediate realized negative stock return; (5) the investigation on the coherence in regimes over time for the six countries shows that some countries show increasing coherence in regimes, like the SWE-CAN and SWE-US pairs, many counties show decreasing co-movements in volatility regimes

(e.g. the SWI-UK and SWI-US); (6) given the evidence of decoupling in coherence of volatility regime over time, future research on the implication of this effect on global portfolio diversification and investment strategy is promising.

In the last essay, we decompose the term premium of the North American CDX investment grade index into a permanent and a stationary component using a Markov switching unobserved component model. We consider an appropriately specified Markov switching unobserved components model as a reliable measure of volatility dynamics of the CDX index spread curve and investigate the presence and significance of both monetary policy adjustments and stock market returns for the US economy over the sample period September 2004 - July 2009. To the best of our knowledge, this is the first direct empirically based evidence that is brought on the evolution of the term premium of the CDS index market and its observed macroeconomic and financial determinants. We contribute to the literature by establishing that: (1) the inversion of the CDX index term premium is induced by sudden changes in the unobserved stationary component, which represents the evolution of the fundamentals underpinning the risk neutral probability of default in the economy; (2) we find strong evidence that the unprecedented monetary policy response from the Fed during the crisis period was effective in reducing market uncertainty and helped to steepen the term structure of the CDX index, thereby mitigating systemic risk concerns; (3) the impact of stock market volatility on flattening the term premium was substantially more robust in the crisis period, and equity returns also make a significant contribution to the CDX term premium over the entire sample period.